

Problems with the president

Rapid-fire decisions on ergonomics, arsenic levels and carbon dioxide emissions indicate that scientific opinion sits low in the pecking order of influence inside the new Bush administration.

It normally takes several months for a new US administration to find its tone, as thousands of mid-level political appointees arrive in Washington to man the levers of government, and relationships are established with other power centres, including the Congress. But a string of regulatory decisions made by George W. Bush and others in recent weeks make it abundantly clear where his administration stands on matters in which scientists would normally play an important advisory role. It stands firmly with the employers and polluters who helped to pay for Bush's singularly unimpressive election victory last November, and damn the scientific evidence.

The first decision, on ergonomics, was instigated by the Republican-controlled Congress, which passed a law, immediately signed by Bush, to debar ergonomic regulations, proposed by the previous administration, on the grounds of their alleged cost (see *Nature* **410**, 292; 2001). Although the action was widely interpreted as a successful attack by business interests on the labour unions, it also inflicts damage on the political prestige of the National Academy of Sciences, whose recent report on the matter was blithely disregarded.

The credibility of Christie Whitman, already seen as marginalized in her role as Bush's administrator of the Environmental Protection Agency (EPA), took a further dip last week when she issued a stalling order on a proposal left over from the Clinton administration that would have sharply reduced the level of arsenic in drinking water (see page 503). A National Academy study on this topic found that the cancer risk at the current level — which dates from 1942 — is extraordinarily high. Whitman is promising a new rule soon, but meanwhile the administration is in the invidious position of arguing that the United States cannot afford the arsenic standard that is already required in Europe and is recommended by the World Health Organization.

The most economically and politically significant of the three decisions came on 13 March, when the administration wrote to four senators informing them that it would renege on its pre-election promise to regulate emissions of carbon dioxide (see *Nature* **410**, 401; 2001). The wording of this letter has produced bemusement among even moderate Republicans, who note that it fails to offer any consistent explanation for its summary rejection of the accumulated scientific evidence that greenhouse-gas emissions are contributing to climate change.

The reversal on carbon dioxide has been accompanied by some shameless dissembling by Bush's supporters over the circumstances in which he originally promised to regulate carbon dioxide emissions on the campaign trail. It is being falsely claimed that the promise was barely noticed at the time it was made, and that its insertion in a campaign speech was a "mistake" made by low-level Bush operatives.

On the contrary, the promise was made as part of a concerted effort to portray Bush as a new kind of conservative, sensitive to ordinary suburban voters' concerns about the environment and, it was even suggested, more likely to make real progress on environmental issues than Al Gore, the green evangelist. But instead of bringing a new pragmatism to the climate-change issue, and perhaps providing American leadership on the control of domestic emissions that could have opened the way to renegotiating the unrealistic targets set under the Kyoto Protocol, Bush has seen fit to capitulate to the coal industry at the first available opportunity.

One price — perhaps intentionally exacted — is the humiliation of Whitman, who spent her first month at the EPA telling everyone who would listen that Bush intended to get serious about climate change (see *Nature* **410**, 133; 2001). But a bigger price will be paid by many others if Bush persists in ignoring what science is telling him. ■

Fossil-fuelled feuds

Palaeontology and local politics make troublesome bedfellows.

An unseemly row between rival palaeontologists in Kenya (see page 508) provides a glimpse of the explosive mix of internecine rivalry between palaeoanthropological research groups fighting for mining rights to rare fossil sites, and the politics of the African countries that are home to some of the world's most precious digs. Clashes have also occurred in Uganda, Ethiopia and elsewhere.

Underlying such disputes is a resurgence in many African countries of a desire to take control of their fossils after decades of safari research by foreign groups who, for reasons both within and beyond their control, have largely failed to build local scientific competence. The foreign groups of palaeoanthropologists established in Africa have little incentive to change the status quo. But what is important is not so much who makes the rules — important for establishing priority to those who have invested long and hard in a particular site, and providing for protocols that best preserve stored fossils and regulate their export — but how they are enforced.

Palaeoanthropologists should spend less time fighting with one another, and pay more attention to their common enemy: the degra-

dation of rare fossil sites — mankind's common heritage. The international palaeontology community must develop methods for impartial arbitration to resolve disputes, do more to encourage ethical conduct, and work with agencies such as UNESCO to protect such sites.

As with any science, palaeontology's future reputation depends on principles of good conduct. Young researchers should accordingly consider commandments recommended by Tim White from the University of California at Berkeley in "A view on the science: physical anthropology at the millennium" (see *American Journal of Physical Anthropology* **113**, 287–292; 2000): "In the field do not think you are going to step out of the vehicle and pick up a hominid within 20 meters on the first day. Do not claim that you found the fossil if the other person did. The truth will eventually come out. Do not purchase fossils. Do not reject papers or grant applications for personal or political reasons. Do not bribe officials. Do not steal another person's site, particularly when that person is a local scholar in a developing country." And, it might be added: to avoid charges of bad behaviour, keep accessible and verifiable records of your activities. ■

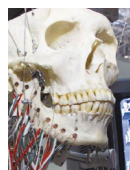




Free range
Publishers come
under fire over
access policies
p502



Vitamin vexation
Experts question
usefulness of
Golden Rice
p503



Profit potential
Japanese researchers
set to gain as patent
restrictions are lifted
p504



Pot boiler
Medical agency
criticized over
cannabis tests
p505

Delays allowed foot-and-mouth epidemic to sweep across Britain

Jim Giles, London

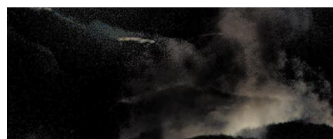
Early delays in tackling food-and-mouth disease undermined the British government's attempts to control the outbreak, according to three epidemiological studies. All of the studies find that the epidemic is now out of control in the United Kingdom.

Animals at a farm near Newcastle in the north of England were almost certainly infected with the disease in early February, but the outbreak was not detected until sheep were moved over 400 kilometres to an abattoir in Essex. By the time full restrictions on the movement of animals were imposed on 23 February, transport of animals infected at the farm and abattoir had spread the disease across the country.

Teams from Imperial College, London, the University of Edinburgh and the government's Veterinary Laboratories Agency (VLA) in Surrey conducted separate studies on the basis of available government data.

"We're still in the exponential stage of the outbreak," says Mark Woolhouse, an epidemiologist at the University of Edinburgh (see Commentary on page 515).

The three studies make different predic-



Out of control: culling is on the increase as foot-and-mouth spreads out across Britain — and Europe.

tions for the epidemic's development in the medium term. The VLA offers the most pessimistic scenario, projecting 4,411 cases by the end of June. At the other end of scale, the Edinburgh team estimates that 918 cases will be recorded by July — 627 cases had been

confirmed by 27 March. Neither team was able to say when they thought the disease would be wiped out.

The government is now trying to limit the spread of the disease by culling all animals within 3 kilometres of infected premises in areas where the virus is widespread — a move that may result in the slaughter of some 600,000 animals.

Far smaller outbreaks have been confirmed in the Republic of Ireland, France and the Netherlands. Dutch vets have got permission from the European Union to vaccinate all animals in a 10-kilometre ring around the outbreaks, although the Dutch government says it will vaccinate only if it cannot slaughter all the infected animals quickly enough.

Despite calling in the Army to help dispose of culled animals, Britain now faces exactly such a backlog, leading some epidemiologists to suggest that vaccination might be useful. It had previously been ruled out because many countries will not accept meat imports from vaccinated areas.

Epidemiologists instead suggest cutting the interval between identifying and slaughtering infected animals. The gap averaged two to three days at the start of the epidemic, but the Imperial group says reducing this by a day could almost halve the total number of cases. ■

Americans perplexed by GM food

Corie Lok, Washington

More than half of Americans say they do not want to eat genetically modified (GM) food, but only one-fifth realize that they are consuming it already. These are some of the findings of a survey by a group set up to encourage a better-informed debate on the issue of transgenic foods.

A quarter of the people surveyed believe that GM foods are unsafe, and slightly more (29%) consider them safe. The remainder are not sure either way, according to a telephone survey of 1,000 adults, performed for the newly launched Pew Initiative on Food and Biotechnology, based in Washington.

However, after being told that more than half of grocery store products in the United States already include GM foods, the percentage of respondents believing the

foods to be safe jumped from 29 to 48.

The survey results "suggest that consumer opinions about safety are not strongly held and may in fact be up for grabs," says Mike Rodemeyer, executive director of the Pew Initiative, a \$12 million public information project supported by the Pew Charitable Trusts.

This change of opinion in response to information about GM foods shows that "mandatory labelling and safety testing will go a long way to improving people's opinions about food biotechnology," says Jane Rissler of the Union of Concerned Scientists, which supports such labelling. The Biotechnology Industry Organization, a trade association, declined to comment on the survey results.

♦ <http://www.pewagbiotech.org>

Publishers challenged over access to papers

Meredith Wadman, Washington

A battle over access to biomedical research papers was no closer to a resolution this week, after two leading journals said they would make their papers freely available on the web within a year of initial publication.

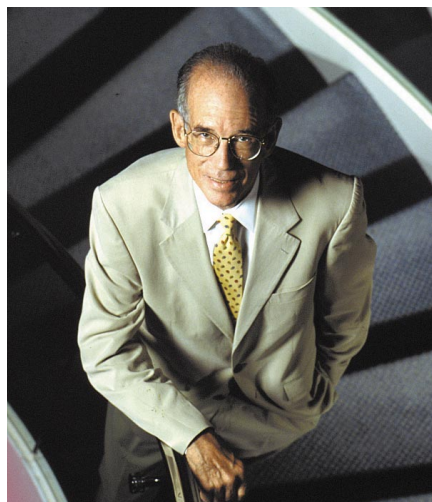
Scientific publishers and some biomedical researchers have been arguing for a couple of years now about the circumstances in which papers should be made available on the Internet. Early last year, the US National Institutes of Health (NIH) introduced a forum called PubMedCentral on which journals can deposit their contents, but few leading journals have chosen to participate.

Now, a group of researchers say they will stop buying, publishing in or reviewing for any journal that refuses to place its research papers in a proposed Public Library of Science (PLS) within six months of their initial publication.

Publishers say that free access could leave journals without enough revenue to support editing or peer review. They have also warned that the proposed PLS might, in effect, bring biomedical publishing under the control of the US government, raising problems about political control and international acceptance.

Science said in an editorial last Friday (23 March) that it will make its content freely available on its website a year after publication, and the *Journal of Cell Biology* said it would do the same after six months.

Donald Kennedy, the editor of *Science*, which is published by the American Association for the Advancement of Science, says of free access: "A lot of people agree that basically this is a good idea. The question is, how can an array of journals do this without exposing



Controlling interests: *Science*'s Donald Kennedy warns of a 'tax supported central monopoly'.

themselves to serious economic risk?"

By 27 March, 12,095 scientists had signed an online petition calling for the establishment of the PLS. The signatories, who include many prominent molecular biologists, pledge to begin their boycott in September. Their leaders include Pat Brown, a geneticist at Stanford University, and Harold Varmus, the former director of the NIH, who is president of Memorial Sloan-Kettering Cancer Center in New York.

The initiative's leaders contend that efficient searching is as important as free access, and so the new policies announced by *Science* and *JCB*, while welcome, still fall short. They say they want to see a competing set of public and private databases, at any one of which scientists could find most of the literature.

But publishers think that the actual impact



Free for all: Harold Varmus believes publishers are holding on to content they should release.

of the initiative would be to place the entire biomedical literature in one public database — probably PubMedCentral, which was launched at the NIH under Varmus. "One has to ask oneself whether a private site is going to arise to compete with an already-established, already-populated, fully-tax-supported central monopoly," says Kennedy. "That doesn't happen very often in the real world."

Varmus responds: "Nobody is arguing for a monopoly here. The only monopoly I see is what seems to be an attempt by journals to hold onto content that at a certain point they should let go."

Some of the loudest complaints about the PLS proposal are coming from scientific societies that rely heavily on journals for income. They say that they will be disproportionately hit by the threatened boycott. "[Commercial scientific publisher] Reed-Elsevier is probably sitting there chortling 'Let the society publishers go out of business,'" says Martin Frank, the executive director of the American Physiological Society, which publishes 14 journals. "They've got deep enough pockets to say: 'To hell with this boycott.'"

Annette Thomas, managing director of the Nature Publishing Group, says that the group welcomes discussion of the issue with researchers. "Many complex issues have been raised and we are currently soliciting feedback from scientists, librarians and other interested parties," she says.

Derk Haank, chief executive officer of Elsevier, says: "We are looking at the issue — but our concern is that there must be an acceptable business model."

But Bruce Alberts, the president of the US National Academy of Sciences, suggests that journals can post free back-content relatively promptly without undue financial risk. "Six months or a year should be safe for any journal," he says.

Butchery lay behind CJD cluster

David Adam, London

Traditional butchery practices are the most likely cause of Britain's first cluster of variant Creutzfeldt-Jakob disease (vCJD) in the Leicestershire village of Queniborough, an official public health report concluded last week. The investigation into the deaths of five people from vCJD, the human form of BSE, also pins down the estimate of the disease's incubation period to between 10 and 16 years.

The inquiry team believes that the disease was probably spread as butchers split open the heads of infected cow carcasses to remove brain tissue, which can contain the BSE agent. This could have allowed the sticky, tainted tissue to be transferred to other cuts of meat from knives, hands, slabs and aprons. Removing

brain tissue was popular during the Second World War as a cheap source of protein, but the practice was relatively rare in Britain during the 1980s and has since been banned.

Four of the five victims ate beef bought from one of two butchers who either slaughtered animals and removed the brains themselves or used a small, local abattoir during the first half of the 1980s, the inquiry found.

But Robert Will, director of the UK National CJD Surveillance Unit, says that the low number of cases involved makes it difficult to make more general predictions from this cluster. "These may be early incubation periods so what really matters is the mean period," he says, adding that other cases of vCJD will almost certainly appear after much longer periods of time.

Critics claim 'sight-saving' rice is over-rated

Nina Schnapp and Quirin Schiermeier, Munich

Advocates of agricultural biotechnology are overselling the usefulness of genetically modified Golden Rice in fighting malnutrition, some experts say. The rice is altered to provide extra vitamin A, deficiencies of which cause blindness and other illnesses.

The sceptics point out that, in the absence of a balanced diet containing fats that can dissolve vitamin A — and so enable the body to absorb it — the widely vaunted health benefits of the rice are likely to elude the poorest people who eat it.

Golden Rice is genetically engineered to contain provitamin A, the precursor of vitamin A. The commercial rights of the rice are held by Syngenta of Basel, Switzerland, which was formed last year from the merger of the agricultural biotechnology arms of AstraZeneca and Novartis.

Syngenta has agreed to deliver the rice licence-free to scientists and subsistence farmers in developing countries, and the first samples were shipped to a rice research institute in the Philippines in January (see *Nature* 409, 551; 2001).

Syngenta has stressed that Golden Rice is only a potential solution to vitamin A deficiency and needs "proper investigation".

But others have made stronger claims for it. They include one of the product's inventors, Ingo Potrykus, a former plant scientist at the Swiss Federal Institute of Technology in Zurich. Potrykus writes in the

forthcoming issue of *In Vitro – Plant* (vol. 37, issue 2) that those opposing use of the rice in developing countries should "be held responsible for the foreseeable unnecessary death and blindness of millions of poor every year".

But some nutrition experts are warning about the product's limitations. "Many children go blind even though they consume sufficient provitamin A," says Francis Reed, a biochemist at King's College London. "This is because dietary lipid is in notoriously short supply in many Third World diets."

Vitamin A is present naturally in foods such as eggs, milk and vegetables. But, like many other nutrients, it is a lipid-soluble substance that can only be absorbed by the body if consumed with sufficient fat or oil.

In countries where the consumption of dietary fat is low, Golden Rice is unlikely to benefit health, says Michael Krawinkel, director of the Institute for Nutritional Science at the University of Giessen in Germany, and an expert on nutrition in poor countries.

Potrykus says he is aware of the problem. "The rice does contain some oil," he says. "But it is unclear whether it is enough to entirely absorb the provitamin A." He agrees that further research is necessary to resolve the issue.

"Its biological usefulness and safety have yet to be examined in animal models," says Krawinkel. Tests involving people suffering from vitamin A deficiency are ethically



Gold reserve: Potrykus, left with colleague Peter Beyer, admits the rice needs further research.

problematic because the deficiency requires immediate medical treatment that could be readily provided by study teams.

Cultivation of Golden Rice will start later this year in India, China and the Philippines, according to Potrykus. He hopes that sufficient rice will be available by the end of the year to carry out systematic research on its properties and modes of consumption.

But Reed questions whether Golden Rice will ever meet the expectations that have been attached to it. "Rather than merely adding vitamin A, the point is to improve poor diets generally," he says. ■

IRRI / ARIEL JAVELLANA

Plans to reduce acceptable arsenic limit put on hold

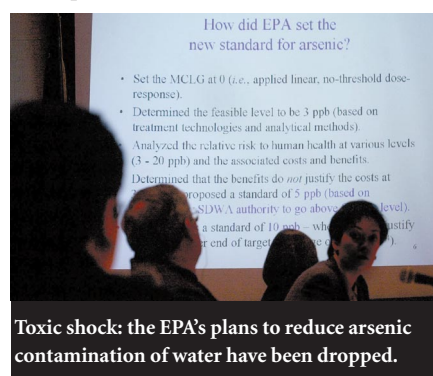
Corie Lok, Washington

The US Environmental Protection Agency (EPA) is halting plans to lower the standard for arsenic in drinking water from 50 parts per billion (ppb) to 10 ppb, citing a lack of scientific evidence to justify the change.

The EPA says it will now seek an independent review of the science behind the lower standard, despite a 1999 study conducted by the National Academy of Sciences (NAS) at the EPA's request. In its report, the NAS recommended that the 50-ppb standard, set in 1942, be lowered "as promptly as possible" to an unspecified number.

A 10 ppb limit on inorganic arsenic is already recommended by the World Health Organization and by a 1998 directive from the European Union.

"Certainly the standard should be less than 50 ppb, but the scientific indicators are unclear as to whether the standard needs to go as low as 10 ppb," said EPA Administrator Christie Whitman in a statement on the cancellation of the 10-ppb standard. She



added that a new standard will quickly be devised to take its place.

But the former head of the EPA's water division, Chuck Fox, who was involved in drafting the 10-ppb rule, says that the available data could support a limit of as little as 5 ppb. "There's no question that there's enough science to justify bringing the standard down to 10 ppb," he says.

The NAS report found that risks at low-level exposures to arsenic had not yet been

directly established, but that extrapolation of existing data suggested a cancer risk of between 1 in 100 and 1 in 1,000 for people who regularly consumed water containing 50 ppb of arsenic.

A fresh study in this month's issue of the *American Journal of Epidemiology* (153, 411–418; 2001) is one of only a few that directly measure risk for arsenic levels around 10 ppb. It found that the risk of urinary tract cancer in Taiwanese farmers was significantly higher when consuming well water containing more than 100 ppb of arsenic compared with those consuming less than 10 ppb. But the increased risk among subjects drinking between 10 and 50 ppb was not statistically significant.

The 10-ppb arsenic rule was issued during the Clinton administration's last week in office, and Whitman considers it may have been a rushed decision. Fox disagrees, saying it is based on more than a decade of scientific and economic analysis. ■

♦ www.nap.edu/books/0309063337/html/index.html

♦ www.aje.oupjournals.org/content/vol153/issue5

AP

Race is on to win Australian funding for big science projects

Peter Pockley, Sydney

The Australian government has opened a competition among groups of researchers for A\$155 million (US\$76 million) to build major scientific facilities.

Several telescope projects and a synchrotron are already in the frame for the funding, which was announced this month by science minister Nick Minchin. But with statements of intent needed within a week of the announcement, and applications due by 11 May, the money is expected to go to projects that are already well planned.

Such proposals include the scheme from a consortium of nine universities for a synchrotron called Boomerang. Physicist John Boldeman, facility director of the Australian Synchrotron Research Program, says that Boomerang would be a 3 giga-electron volt ring with nine beamlines.

The location for the synchrotron has yet to be decided, but three state governments, Queensland, Victoria and New South Wales, say that with the help of local industry they could raise the matching funds the government is demanding of successful project applicants.

Astronomers, meanwhile, are proposing several options for investment — although they may struggle to raise the matching funds. Rachel Webster, a physicist at the University of Melbourne and chair of the National Committee for Astronomy, says that the top priority is to double Australian time on 8-metre telescopes, either by greater participation in the Gemini project or by joining the European Southern Observatory.

But astronomers at the University of New South Wales want to build a 2-metre, wide-field infrared telescope in Australian Antarctic Territory. A fourth telescope plan would build a 6.5-metre, wide-field optical telescope in Chile. ■



Cold interest: Australian and French astronomers site-testing in Antarctica.

Japan's academics get green light to make their fortunes

David Cyranoski, Tokyo

From now on, the sky's the limit for academic entrepreneurs in Japan. The government has removed a ¥6 million (US\$50,000) ceiling on the amount that researchers at public universities are allowed to earn from government-held patents based on their work.

Removing the limit is the latest of several steps designed to encourage academics to pursue the industrial applications of their work. The Japanese government holds the patent rights for much of the research done at public universities. And the number of patents granted on university-based science in Japan has been growing steadily in recent years (see graph), having jumped almost fivefold over the past 10 years. But many of these lie unused.

Restrictions preventing technology licensing offices from dealing with government-held patents were also removed last week. Seventeen such offices have opened in affiliation with Japanese universities since a 1998 law permitted them to be set up.

These policy changes could persuade researchers to be more active in pursuing applications for their work, says Takafumi Yamamoto, president of the Center for Advanced Science and Technology Incubation (CASTI), a licensing office at the University of Tokyo. "The government is no longer saying that it's bad to make money," he says.

CASTI's recent success has paved the way for an additional office funded by Tokyo University researchers themselves. The Advanced Science and Technology Enterprise Corporation, which opens in April, will focus on developing basic science discoveries into marketable products.

Researchers at the Institute of Physical and Chemical Research (RIKEN) are also catching on: 16 researchers there have invested in a venture business that will next month start to patent and license ideas from RIKEN.

The current burst of activity is regarded by some as a sign of how far Japan lags behind other countries in encouraging money-making by academics. "The really surprising thing about the ¥6 million limitation is that it was still on the books at all," says Yukio Kato, manager of Waseda University's Intellectual Property Center in Tokyo.

Other obstacles to getting and using patents remain. Whereas a provisional application procedure allows US researchers, for example, to file patents at a minimal cost, Japanese researchers have to spend ¥300,000–400,000 in filing an initial applica-

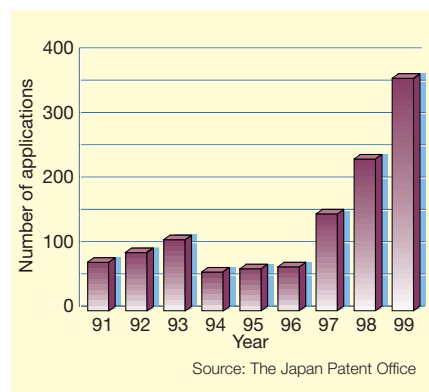


Profitable tool of the trade: a mastication robot developed by researchers at Waseda University.

tion. And any new business is legally required to demonstrate that it has at least ¥10 million in capital. "This makes it hard for people to take a chance by starting on a small scale with a new technology," says Yoichi Okabe, director of the Research Center for Advanced Science and Technology at the University of Tokyo.

Other restrictions prohibit interactions between academia and industry. As public servants, for example, university researchers are not allowed to use their university e-mail for business matters. There are also strict restrictions on holding stocks and taking up stock options.

Some observers say that the system can only be substantially reformed after a proposed increase in university autonomy, tentatively scheduled for 2003 (see *Nature* 401, 198; 1999). ■



A patent increase: changes in the number of patent applications from Japan's universities.

UNIV. TOKYO

Sonar system offered special dispensation

Mark Schroppe

The US government is proposing to waive environmental laws so as to enable the Navy to deploy a controversial sonar system for submarine detection. The sonar has been the subject of intense debate because of its potentially adverse effects on marine mammals.

The Surveillance Towed Array Sensor System, with a Low Frequency Active upgrade, listens for reflections bouncing off submarines in response to bursts of sound emitted by the system. The Navy says the system is needed because modern submarines are too quiet to be detected by passive systems. The sonar has so far been fitted on only one ship, but the Navy has funding for a second and hopes to build two more to obtain global coverage.

Tests on the system were classified information until about six years ago, when environmentalists learned about it. They threatened the Navy with lawsuits based on concerns that it might cause hearing loss or disorientation in marine mammals. The Navy responded by agreeing to assess the system's environmental effects. It also supported additional research on the sonar's effects on the animals.

The research found that the system might significantly affect the mammals by, among other things, altering their migration patterns. Two recent whale strandings have also been linked to sonar use, although the Navy says these events involved factors, such as proximity to land, that would not apply to their plans for deploying the system.

Based on the assessment and related research, the National Marine Fisheries Service (NMFS), part of the National Oceanic and Atmospheric Administration, has just released a draft of a new rule that would give the Navy a five-year exemption from the Marine Mammal Protection Act. The act prohibits the harassment or killing of marine mammals by US entities in or outside US waters.

The rule outlines mitigation measures that the Navy and the NMFS say will ensure that the system has little effect on marine mammal populations. They include a promise not to use the system within 20 kilometres of the coast, to shut it down whenever mammals are detected within one kilometre, and to stay out of regions where mammal aggregations are common.

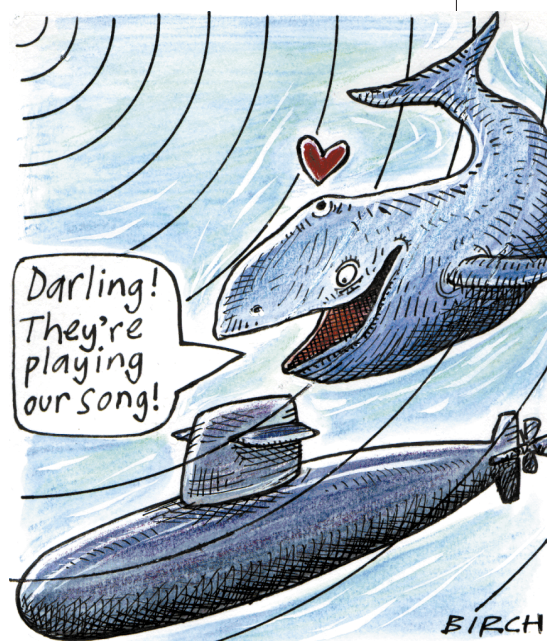
Environmentalists and some academics say that, although there are insufficient data to be certain about the system's effects on marine life, the mitigation measures probably do not go far enough. "I simply do not believe that the Navy has addressed the most fundamental problems," says Joel Reynolds, director of marine mammal protection at the

Natural Resources Defense Council, an environmental group.

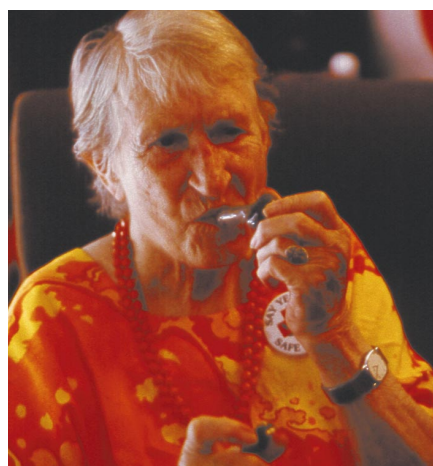
Jonathan Gordon, who studies marine mammal acoustics at the University of St Andrews in Scotland, expresses satisfaction that the effects of the military technology have been assessed at all. He points out that France and Britain are developing comparable systems without any public scrutiny. But he remains concerned that the US proposal is based on inadequate information.

Ken Hollingshead, a fisheries biologist at the NMFS, says that the Navy recognizes the need for better information, and will continue to seek it. But Roger Gentry, coordinator of the NMFS acoustics team, says he believes there is enough information available to say with reasonable certainty that the system will cause no significant harm.

Public comment is being invited on the proposed rule, and a final decision will be announced later in the year.



Regulator rebuked over cannabis



Drug culture: therapeutic uses for cannabis may be held up by demands for toxicity data.

Alison Abbott

The UK Medicines Control Agency (MCA) has come under fire for insisting on full toxicology testing of cannabis extracts in clinical trials, allegedly ignoring accumulated knowledge of the drug's safety.

The Science and Technology Committee of the House of Lords charges in its latest report that the MCA is "not approaching the question of licensing cannabis-based medicines in a properly balanced way".

A 1998 report from the same committee recommended that doctors be allowed to prescribe cannabis-based medicines. Some patients in Britain with diseases such as multiple sclerosis already treat themselves

with cannabis, although it is illegal to do so.

Since the 1998 report, several clinical trials of cannabis have begun in Britain. These include two funded by the Medical Research Council that are comparing an oil-based cannabis extract with the major active component of cannabis.

Trials with pure extracts of different plant varieties of marijuana are also being conducted by the Porton Down-based company GW Pharmaceuticals. Assuming these trials are successful, a product could be available by 2003. But a further delay of two years would be inevitable if long-term toxicology studies are needed, says Geoffrey Guy, the company chairman. He adds that Canadian authorities are not asking for extensive toxicology testing.

Leslie Iversen, professor of pharmacology at the University of Oxford and a scientific advisor to the Lords committee, says the MCA is treating cannabis as if it were an entirely new drug. The agency should be much more flexible in its demands for toxicological data, he says, as experience has shown it to be safe.

"Cannabis was in the pharmacopoeia until the 1940s," adds Lord Perry of Walton, who chaired the Lords' committee. "Asking for two years more testing, as if it were a new compound, is carrying things a bit far." Since the report was written, the MCA has told the Lords committee that it is prepared to review its position.

♦ <http://www.parliament.the-stationery-office.co.uk/pa/ld200001/ldselect/ldscitech/50/5001.htm>

REX FEATURES

Major expansion planned for University of Cambridge

London The University of Cambridge has announced ambitious plans to increase the number of its faculty members by 3,000 and student enrolment by 5,000 over the next 25 years.

Three new colleges are planned. They would be built to the north-west of the city on university land that is currently used for agriculture teaching and research.

University vice-chancellor Sir Alec Broers says the extension is needed to maintain the university's standing and its contribution to the regional economy. But the use of open land for the development is set to attract local opposition.

♦ <http://www.admin.cam.ac.uk/univ/northwest/>

Patent dispute ends in settlement

San Diego The California-based company Affymetrix and Oxford Gene Technology (OGT) of the United Kingdom have settled a long-running legal dispute over the development and patenting of microarray technology.

In a case due to be heard in the Delaware

Federal Court in the United States, OGT had claimed that Affymetrix infringed its patents by developing its GeneChip technology, used to determine which genes in a cell are switched on. Affymetrix had responded by claiming that OGT's patents were invalid (see *Nature* 404, 697; 2000).

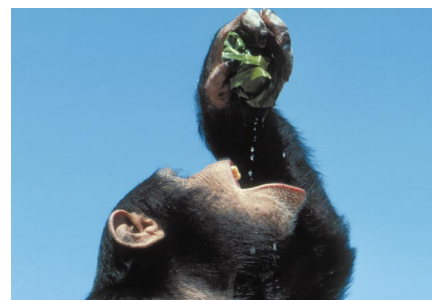
Both organizations have now dismissed their respective US lawsuits, as well as a series of infringement actions and patent challenges pending in the United Kingdom and Europe. Financial details of the settlement were not disclosed.

Campaign against chimp research centre

London The Biomedical Primate Research Centre, a non-profit foundation specializing in the use of non-human primates for research into human disease, is to be targeted for closure by a coalition of European animal welfare organizations.

The centre, based in Rijswijk in the Netherlands, uses a stock of over 1,000 primates for studies in transplantation, immunology and drug development. It conducts collaborative experiments with industrial partners and also supplies animals to other European laboratories.

The Coalition to End Experiments on Chimpanzees in Europe, which has the



backing of primatologist Jane Goodall, alleges that animals at the centre are kept in cramped conditions, and claims they are no longer needed for many of the experiments they were originally bred for.

♦ <http://www.hollandbiotechnology.nl/FIRM/BiomedicalPrimate/BPRC.html>

Letter laments exclusion from Nobel prize

London More than 250 neuroscientists have published an open letter saying that Oleh Hornykiewicz, a neurologist at Vienna University Medical School, should have been included among the winners of last year's Nobel Prize for Medicine.

In 1960, Hornykiewicz showed that the brains of patients with Parkinson's disease had low levels of the neurotransmitter dopamine. One year later he reported the

first treatment of Parkinson's with levodopa — still considered to be the most effective drug for treating the disease.

The letter was published in the current issue of *Parkinsonism and Related Disorders*. But the authors do not expect a response: the Nobel Foundation, the body responsible for awarding the prize, does not publicly discuss the selection committee's choices.

Japan sets genetic modification limits

Tokyo The Japanese government says it will place a 5% upper limit on the amount of transgenic corn that can be contained in a product labelled as 'non-genetically modified'.

The 5% level matches rules set two years ago which limit the transgenic content of soybean products. Products missing the 5% mark will carry an 'unclassified' label. Consumer groups are unlikely to be happy with the ruling: the European Union, they point out, requires that products labelled as non-genetically modified should contain less than 1% of transgenic food.

Japan's Ministry of Agriculture, Forestry and Fisheries says that the policy is adequate because any product that carries potential health risks will be banned outright.

Astrobiology Institute takes new partners

San Diego Four new institutions have joined NASA's Astrobiology Institute, taking its total number of research partners to 17.

The institute's new members are the Jet Propulsion Laboratory in California, the University of Washington in Seattle, Michigan State University and the University of Rhode Island. Each receives a \$4 million NASA grant spread over five years.

Based at the Ames Research Center near San Francisco, the institute studies diversity of life in extreme environments and develops models to search for habitable planets outside our Solar System.

► <http://nai.arc.nasa.gov>

New head named for EU research centre

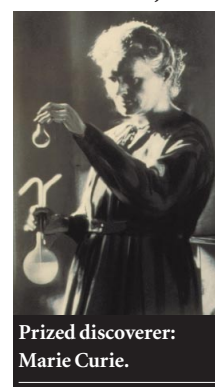
Munich The European Commission's Joint Research Centre (JRC), a conglomeration of eight research and service institutes, has appointed Barry McSweeney as its new director-general. McSweeney currently directs the JRC's Institute for Health and Consumer Protection in Ispra, Italy.

The JRC, which conducts research to assist the European Commission with science-based decisions, has been criticized

for over-diversifying its activities and running ineffective programmes. McSweeney admits that the JRC has been "thinking in an apologizing and defensive manner for too long". He says that its research laboratories should focus on consumer-sensitive areas such as food safety, the release of genetically modified organisms and the safety of chemicals.

Sweden plans centenary celebration for Nobels

London Erwin Schrödinger, Alexander Fleming and Marie Curie are among the scientists featured in the first major exhibition at Stockholm's Nobel Museum. The exhibition, "Cultures of Creativity", uses



Prized discoverer:
Marie Curie.

film and interactive exhibits to examine the inspiration behind the work of 30 Nobel laureates from different fields. It is timed to coincide with this year's centenary of the establishment of the prize.

► <http://www.nobel.se/nobel/nobelmuseum/index.html>

SPL

The battle of Tugen Hills

Behind claims that the oldest human ancestor has been discovered lies a bitter row over access to Kenyan field sites — pitching the bones' discoverers against some of palaeoanthropology's biggest names. Declan Butler reports.

A year is a long time in Kenyan palaeoanthropology. In March 2000, Martin Pickford, a Kenyan-born researcher of British extraction, was arrested on charges of collecting without a research permit and spent five days in a Kenyan jail. Now he is the proud co-author of a paper¹ describing *Orrorin tugenensis*, or the 'Millennium man' — a 6-million-year-old find that Pickford and his colleagues maintain is the oldest known human ancestor.

This claim has sparked intense debate — *Orrorin's* age is not in doubt, but some researchers question whether it is a member of the human lineage (see page 526). And the scientific controversy is only the beginning. No sooner had Pickford and his colleagues announced the find, at a press conference held in Nairobi last December, than Andrew Hill, chair of anthropology at Yale University in Connecticut, accused Pickford of encroaching upon his site in the Tugen Hills, part of Kenya's Baringo region. Pickford denies this, and alleges that Richard Leakey, former director-general of the National Museums of Kenya, and until recently head of the Kenyan civil service, incited last year's arrest. Pickford is now suing him for damages.

The controversy provides a glimpse of the bitter rivalries that can flare up between researchers competing for rights to study precious fossil sites, and how these rivalries can become intertwined with the politics of host countries. Although it is not an isolated case, the fact that the Tugen Hills row involves such a prominent figure as Leakey means that it will draw attention to this unseemly side of



Remains of the day: Millennium man's upper arm (right) is just one of the bones unearthed by Martin Pickford and Brigitte Senut.

palaeoanthropological research.

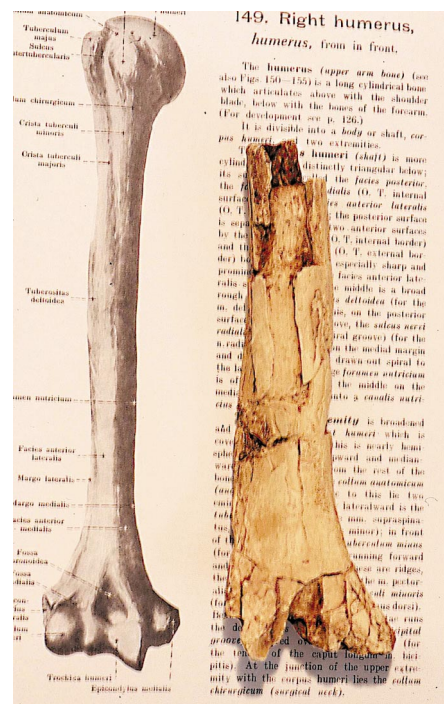
The expedition that yielded the remains of Millennium man is a joint venture between Pickford's institution, the Collège de France in Paris, and the Community Museums of Kenya, a non-governmental organization established in 1997. The project is led by Pickford and Brigitte Senut of the Paris Museum of Natural History.

Permission to dig

Researchers wishing to carry out palaeontological research in Kenya must obtain a government permit, and be affiliated with an officially sanctioned Kenyan research organization. Traditionally, this role has been fulfilled by the National Museums. Some scientists allege that this has, over the decades, enabled the Leakey family to exercise reasonably tight control over who is allowed to conduct palaeoanthropological research in Kenya.

Louis Leakey in the 1930s established the dynasty with his wife Mary, also a renowned palaeoanthropologist. Their son Richard has continued the line with his wife Meave, who currently heads the National Museums' palaeontology division.

Pickford has long claimed to be a victim of this concentration of power. He worked in Kenya for his PhD in the 1970s, discovering a molar tooth² that he argues comes from another specimen of *Orrorin*¹. But Pickford's research in Kenya came to a halt in July 1985,



when Richard Leakey barred him from the National Museums following allegations that Pickford had stolen field notes from the museums. Pickford claimed that he was falsely accused, and that he had been denied due process.

Leakey, who earlier this week resigned as head of the Kenyan civil service, stepped down as director-general of the National Museums in 1989. But Pickford did not return to Kenya until 1998, after finding an ally in Andrew Kiptoon, who was then

minister for research and technology. In a letter dated 3 June 1998, Kiptoon invited Pickford to join with Eustace Gitonga, director of the Community Museums, in developing the archaeological and palaeontological potential of the Tugen Hills, adding that it would "not be difficult to obtain the necessary clearance, work permit, and other approvals".

Bones of contention

From here on, the story becomes almost Kafkaesque. A research permit was granted to Pickford on 30 October 1998. But Hill points to a letter signed by J. E. Ekirapa for the permanent secretary of provincial administration, in the office of President Daniel arap Moi, dated 2 November 1998, cancelling the permit on the grounds that the application had not been properly submitted. It states: "Any research being undertaken on the strength of the cancelled permit is illegal and may lead to further legal proceedings." This letter is also referred to in correspondence dated 21 April 1999 from the same office to the permanent secretary at the then Ministry of Home Affairs, National Heritage, Culture and Social Services, complaining that Pickford's excavations are "illegal and unauthorized".

In March 2000, matters came to a head. Police and National Museums officials raided the Community Museums' Nairobi headquarters and at the same time intercepted Pickford in the field. He was charged with illegal excavation and imprisoned. But proceedings against Pickford came to an abrupt end in mid-April, when Kenya's Attorney-General directed the prosecution to drop the case.

Pickford and the Community Museums are now suing Leakey, the governors of the National Museums, and the Attorney-General, alleging unlawful arrest, false imprisonment and malicious harassment.

In court documents, Pickford claims the November 1998 letter was concocted to disrupt his work. He told *Nature* that he did not receive the letter until November 1999, and argues that it contains several irregularities. But Ekirapa asserts that the letter is genuine. Documents have been filed for the defence that deny all of the charges.

Pickford's dossier of evidence includes a letter dated 14 March 2000 from Leakey to George Abungu, the current director-general of the National Museums³. In this letter, Leakey notes that Pickford is collecting fossils and suggests "that you urgently get assistance from the Director of CID and that you send an officer plus someone from [the National Museums] to intercept Pickford. His possessions should be thoroughly searched and any fossils should be confiscated ... A search should also be made of his premises in Nairobi."

Abungu confirms that he received Leakey's letter, and that National Museums officials were present when Pickford was arrested. But when contacted by *Nature*, Abungu expressed sympathy for Pickford. "It appears he has a research permit," says Abungu. "Pickford has definitely been a victim for a long time." Abungu also supports the right of the Community Museums to provide affiliations for visiting researchers. "The laws have been too sympathetic to the National Museums," he says.

Leakey did not respond to requests to be interviewed for this article, and has not commented publicly on Pickford's allegations. But Hill is trading accusations with Pickford and his colleagues. In a short news story published in *Science*⁴ in December, he described Pickford's presence in the Tugen Hills as "highly irregular". That article also recounted the story of the withdrawal of Pickford's permit, prompting angry responses from both Pickford⁵ and Gitonga⁶.

Irrespective of the dispute over the validity of Pickford's permit, Hill claims that the French team encroached upon a site in which the Baringo Paleontological Research Project, a joint venture between his Yale group and the National Museums, had operated since 1980. "We did a very carefully controlled excavation there, and Pickford

came and dug trenches through the whole site," he says.

Senut says that the French team made enquiries with local authorities to see if Hill was still working in the Tugen Hills before beginning the dig. "It appeared that there was no permit issued for the area," she says. "It appeared also that Dr Hill had not put his feet in the area since 1993." But Hill says he has visited the Tugen Hills every year except one over the past decade, and faxed copies of his permits and other supporting documents to *Nature*. His claims are supported by John Kingston, an anthropologist at Emory University in Atlanta, Georgia. "The last time I was there was in June 1999, and Andrew was there for two months," he says.

Several palaeoanthropologists contacted by *Nature* declined to comment on such an incendiary row, which seems to be fuelled by deep personal animosities. Pickford certainly has antagonized his opponents: in 1995, he co-authored with Gitonga a book called *Richard E. Leakey: Master of Deceit*, which launched a ferocious attack on Leakey and several other leading palaeoanthropologists. Gitonga has also clashed with Leakey — who sacked him from his post of exhibits director at the National Museums.

"It is very difficult to know who is telling the truth," says one US researcher, who asked to remain anonymous. But as the accusations continue to fly, the public image of palaeoanthropology could be the main loser.

Declan Butler is *Nature's* European correspondent.

1. Senut, B. et al. *C. R. Acad. Sci.* **332**, 137–144 (2001).
2. Pickford, M. *Nature* **256**, 279–284 (1975).
3. Fléaux, R. *Sciences et Avenir* No. 648, 18–19 (2001).
4. *Science* **290**, 2065 (2000).
5. Pickford, M. *Science* **291**, 986 (2001).
6. Gitonga, E. *Science* **291**, 986 (2001).

Richard Leakey is not commenting on the row that has erupted over the fossil digs in the Baringo region of Kenya.



Think like a bee

Are silicon circuits that mimic the nervous systems of insects and other animals the future of computing?

Jim Giles considers the merits of neuromorphic engineering.

Rodney Douglas is very impressed with the cognitive powers of the honey bee. “Just look at what it does. It learns how to fly, it can navigate, it has rather sophisticated pattern recognition and some level of communication,” he says. “But it has very grubby sensors. Have you seen the world through a bee’s eye? It’s horrible.”

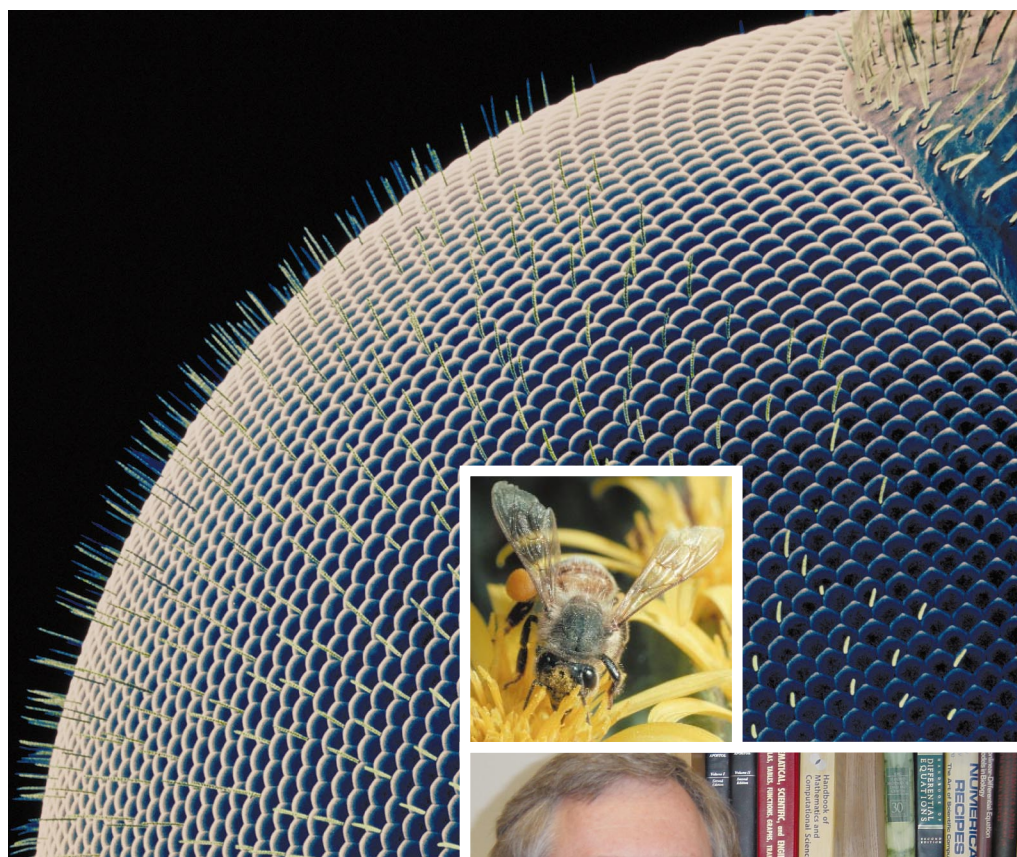
If Douglas were a zoologist, such enthusiasm would not be surprising. But his team at the Institute of Neuroinformatics in Zurich does not study animals — it makes silicon circuits. Douglas is a leading figure in the field of neuromorphic engineering, which attempts to create microelectronic circuitry that mimics the way animal brains work.

Digital computers are extremely good at precise number-crunching. But animals, argue neuromorphics engineers, have evolved brains and senses that let them interact efficiently with the messiness of the real world. So why not develop devices that combine the best aspects of both? “Stupid-looking organisms are doing amazing computation,” says Rahul Sarpeshkar, an electronics engineer at the Massachusetts Institute of Technology. “We need to replicate this in electronics.”

Mind mimics

Attempts to mimic animal brains using artificial neural networks date back to the 1940s. The basic idea is to link together individual processing units — the artificial neurons — that can integrate signals from other units in the network and send signals on to another group of units. Each of the units can only process a limited amount of information, but combined into a network they should, in theory, mimic some of the functions of a real brain.

Neural networks really began to take off in



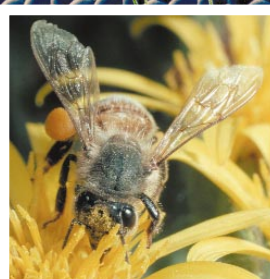
Fresh view: studies of insect eyes have helped Rodney Douglas to produce artificial retinas.

the early 1980s, when researchers developed sophisticated learning ‘rules’ that allowed them to ‘train’ the networks to recognize specific patterns^{1,2}. At that time, programmers were finding it difficult to make digital computers perform pattern recognition.

Signatures, for example, can be a problem. Every time we sign our names, we create patterns that are slightly different from each other. For an automated signature recognition system to work, it must recognize these different patterns as examples of the same signature. This is difficult to do using digital computers, but artificial neural networks can be taught the technique. Different examples of someone’s signature are presented to the input neurons, while the output neurons are set to keep giving the same output. By selectively strengthening connections between the output units and the various intermediary neurons triggered in response to different versions of the same person’s signature, networks can learn to identify signatures.

But building such a neural network using silicon is not easy. Strengthening the connections between individual neurons in the network is done by changing and maintaining the amount of charge stored at the junctions between them. For the pioneers of neuromorphic engineering, this was a big problem: every time the networks were switched off, the charges were lost and the networks ‘forgot’ the tasks they had learned.

Conventional computers do not suffer from such a problem, so instead of building networks from silicon, the researchers simulated them using software, and saved their trained states in a computer’s digital memo-



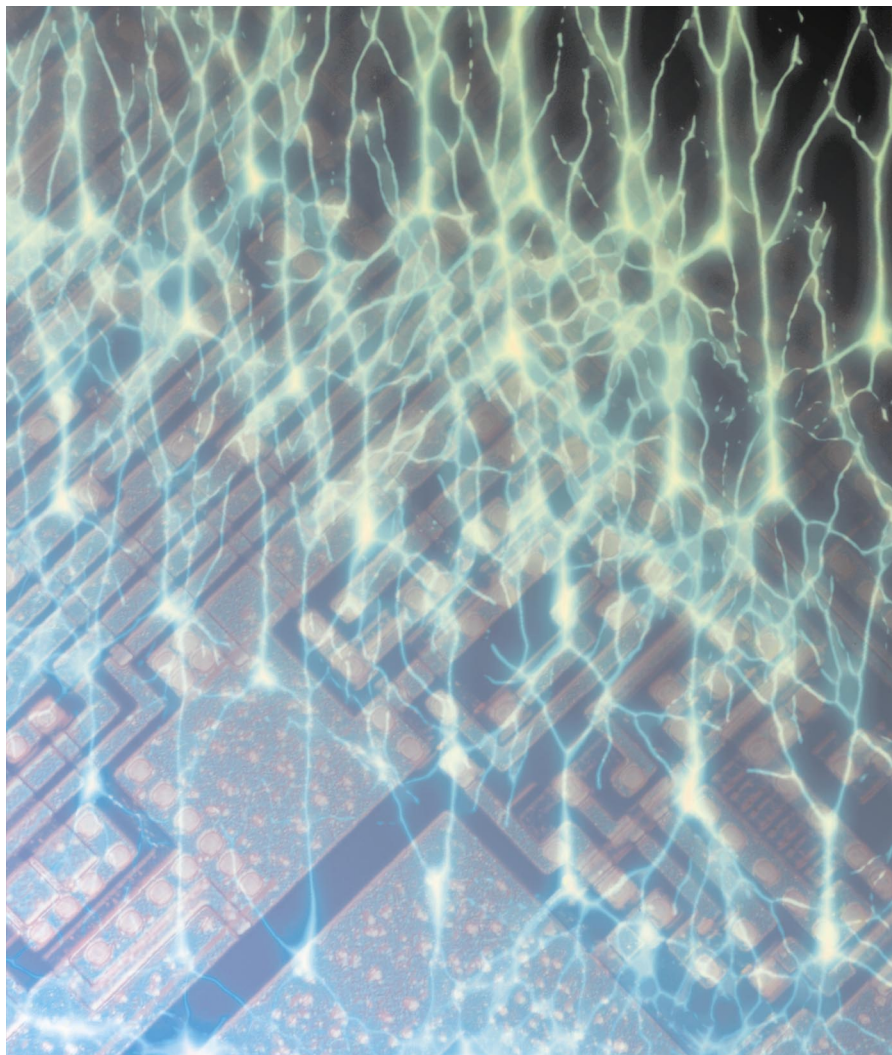
ry. Over the past three decades, simulated neural networks — some more brain-like than others — have been studied extensively. They have shed light on how real brains work and spawned a host of useful applications.

Faking it

But simulated networks deviate from the true spirit of neuromorphic engineering. In biology, brains and sensory organs work quickly and at low power. Simulations, however, are slow and need power-hungry computers. But thanks to research conducted from the late 1980s by Carver Mead and his team at the California Institute of Technology in Pasadena, genuine neuromorphic circuits are now being worked on by about a dozen research groups worldwide.

Mead was interested in creating silicon circuits based on analog electronics. In digital computer chips, problems are solved using strict algorithms. Numbers are represented in binary code — with ‘0’ and ‘1’ corresponding to distinct voltages — and a central ‘clock’ regulates how information is sent around the chip.

Analog circuitry seems chaotic by comparison. A range of voltages is used to represent different numbers, and signals flow



Thought provoking: can the adaptable attributes of networks of nerve cells (green) be married to the precise number-crunching of digital chips to produce a new generation of silicon circuits?

between different parts of the circuit without any central control. Precise algorithms are impossible to implement. Instead, the circuit's architecture is designed so that the 'natural' flow of signals produces useful processing — in much the same way that animal brains work.

One of Mead's biggest contributions was the invention of floating-gate analog structures³, devices that could reliably store charge for long periods of time. This helped to solve the memory problem, and paved the way for neuromorphic devices that simulate the function of the retina, the light-receptive layer at the back of our eyes.

The retina is far more than just a collection of photoreceptor cells. It performs computations, processing information to accentuate the edges of objects, called 'edge extraction', and adjusting the 'gain', or amplification of the signal, to compensate for bright or dark conditions. Powerful digital machines can replicate this 'preprocessing', but nervous systems do so using simple, low-power analog circuitry.

Mead wanted to replicate this simplicity,

and teamed up with the late Caltech biologist Misha Mahowald. The pair's original retina⁴ has since been improved by the invention of better photoreceptors⁵ and more sophisticated circuit design⁶. But the basic philosophy remains the same. Silicon retinas consist of photoreceptor arrays, in which each receptor is connected to its neighbours. A network of resistors, amplifiers and other devices allows the signals to flow between the receptors in real time. Designing this circuitry so that it physically mimics the networks in real retinas is impossible — the number of cells and connections between them is far too great. Instead, neuromorphics engineers work out how retinal networks do their pre-processing and then they design simpler analog networks to do the same job. Modern versions, such as the devices produced by Douglas and his Zurich colleagues⁷, now perform edge extraction and gain adjustment much as biological retinas do.

In 1992, Mahowald added another item to the tool-box of neuromorphic engineering in the shape of address-event-representation (AER)⁸ — a method that allows chips to com-



Carver Mead (seated, centre) and his Caltech team have pioneered chip-based neural networks.

municate with each other. Although neuromorphic chips often contain hundreds of interlinked artificial neurons, there is a limit to the number of connections a chip can make to the outside world. When two neuromorphic chips are connected together it is impossible to link all the individual neurons directly. In AER, incoming and outgoing signals associated with specific units are sent via a central 'bus'. Connections within chips are much easier to manage, so the bus can talk directly to individual units using their 'address' — a number that uniquely identifies every unit.

All-seeing mouse

Neuromorphics engineers can even point to a commercially successful product: the optical computer mouse, introduced in 1994 by Logitech of Fremont, California. Unlike a conventional mouse, which tracks movement using a ball in its base, the Logitech device detects changes in its position using a visual sensor that monitors movement by 'looking' at the desk below. The digital approach to such a problem involves comparing sequential snapshots of the scene as the sensor moves along — much too computationally demanding for a cheap device. Instead, Logitech commercialized a neuromorphic chip developed by André van Schaik, then at the Swiss Center for Electronics and Microtechnology in Zurich. Van Schaik had used his knowledge of a fly's visual system to design a cheap, low-power device⁹.

"The fly's brain compares the change in intensity at one photoreceptor with delayed versions from neighbouring receptors," says van Schaik, now at the University of Sydney in Australia. By analysing the output from neighbouring photoreceptors, cells in a fly's eye make estimates of the speed and direction of the moving objects it sees. But individual 'motion-detecting' cells give only crude estimates of these quantities. More accurate estimates of movement are generated when cells in the later stages of processing compare the output of the different motion-detecting cells. The optical mouse takes a similar approach.

Although relatively simple, Logitech's

► mouse shows how analog neuromorphic devices can be combined with conventional digital computers, melding the latter's high-precision number-crunching with the fast, lower-power pattern analysis that animal brains have evolved to do. But if neuromorphic devices have so much potential, why are there only hundreds, rather than thousands, of researchers working in the field?

One reason is that analog computing has been a victim of the relentless improvements in digital-chip technology. Chip manufacturers are reluctant to invest in an analog speech-processing device, for example, when constantly improving computing power offers new ways of approaching the problem digitally.

Nervous niches

Given this dynamic, neuromorphics engineers are concentrating on niche applications in which the advantages of biologically inspired computing cannot be ignored. For researchers working on mobile devices that must function autonomously, for example, the energy efficiency of neuromorphic devices is appealing.

This is the logic behind using neuromorphic technology to produce 'bionic' implants. Current cochlear implants, used to restore hearing to some congenitally deaf people, contain mechanical versions of the hair cells that sense incoming sound waves. Artificial cochleae are relatively crude, using between 10 and 20 electrodes to simulate the input of 30,000 or so hair cells into the auditory nerve, but the results can be impressive. In the best cases, the implants allow their wearers to conduct telephone conversations.

But current devices use a bulky and power-hungry digital-signal processor that has to be worn externally. And the implant itself also needs recharging every few weeks, when the wearer is forced to sit next to a charging station for several hours.

Toumaz Technologies, a spin-off from Imperial College in London, now aims to produce a smaller, lower-power analog version of the digital-signal processor. Chris Toumazou, the electronics engineer behind the project, says the device, including both new, low-power electrodes and his processor, will fit within the ear and will only need recharging once a year. He hopes to begin clinical trials before the end of the year.

Toumazou is also starting work on turning the analog retinas pioneered by Mead and Mahowald into practical medical implants. This will be more difficult than developing cochlear implants because pre-processing in the retina is much more complex, and many more nerves are involved — there are 1 million fibres in each of our two optic nerves, compared with 30,000 in each auditory nerve.

Meanwhile, Ralph Etienne-Cummings, an electronics engineer based at Johns Hop-

Chips offer insights into vision

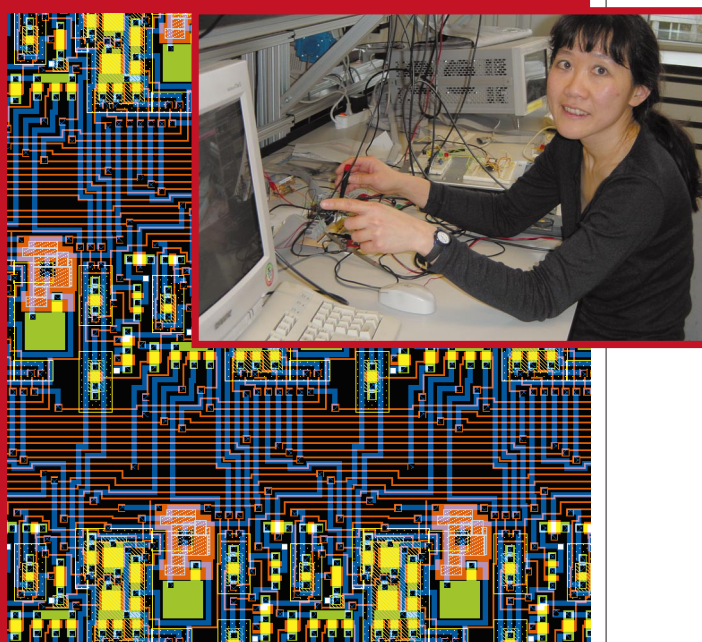
kins University in Baltimore, is working on a circuit that will mimic the way the human spinal cord regulates muscle contraction in the legs during walking. Although we are not consciously aware of it, walking requires sophisticated and continuous real-time computation. Our spinal cord integrates information about our balance and leg positions to calculate the right set of muscle contractions. Etienne-Cummings has teamed up with Iguana Robotics of Mahomet, Illinois, to create a chip that might one day be implanted into the spinal cords of paraplegics to help them walk again.

Even the strongest enthusiasts for neuromorphic engineering accept that the current number of products — one computer mouse — is not that impressive. But a range of successful biological implants would be a differ-

ent matter. "It's been a bumpy road," says Andreas Andreou, a collaborator of Toumazou's based at Johns Hopkins, "but things are now looking very exciting."

Neuromorphics engineers are professional plagiarists, freely borrowing their ideas from nature's bag of computational tricks. But the flow is not one way. By building silicon models of animal brains, researchers can also learn about biology. Many neuromorphics engineers pay lip service to this idea, but researchers at the Institute of Neuroinformatics in Zurich are putting it to work. Neuroscientist Kevan Martin and electronics engineer Shih-Chii Liu are collaborating on a project to investigate how our brains process visual images.

The project centres on a device made by Liu called the cortical chip, an analog silicon circuit based on part of the human visual system. Our brains analyse images using a series of different areas of the cerebral cortex. The first of these, known as 'V1', performs low-level processing such as detecting motion and the orientation of lines. Liu's chip attempts to model two of V1's six layers. It uses a network of artificial neurons connected to mimic the links between layers 4 and 6. Layer 4 is interesting because it receives many of the inputs to V1. To make the chip as realistic as possible, the input to the chip is also taken from biology. Martin has recorded the activity of neurons in part of the cat brain involved in sending signals to



Visual clues: Shih-Chii Liu has designed an analog chip that models some of the brain processes involved in image analysis.

layer 4 of V1. Liu feeds these recordings straight into her chip.

Liu can adjust the connections and properties of the artificial neurons to see how they affect the performance of the whole network. By tweaking specific parameters — such as the number of links between layers 4 and 6 — Liu generates insights that Martin can relate to neurophysiological studies. "The number of questions we have to answer is overwhelming," says Martin.

For example, Martin is interested in how different

neurons in layer 4 form specific 'receptive fields' — in other words, why each responds to a different area of space within the eye's field of view. Each neuron's receptive field is thought to be controlled by inputs from another brain area known as the thalamus. But Martin suspects that there is more to it than that. "Fifty per cent of all connections to layer 4 neurons come from layer 6," he points out. "So what does layer 6 do? With this circuit, we can adjust the layer 6 synapses and use real data to find out."

ent matter. "It's been a bumpy road," says Andreas Andreou, a collaborator of Toumazou's based at Johns Hopkins, "but things are now looking very exciting."

Jim Giles is Nature's assistant News and Features editor.

1. Hopfield, J. J. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
2. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Nature* **323**, 533–536 (1986).
3. Mead, C. *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, MA, 1989).
4. Mahowald, M. *An Analog VLSI System for Stereoscopic Vision* (Kluwer Academic, Boston, 1994).
5. Delbruck, T. & Mead, C. A. *CNS Memo #30* (Computation and Neural Systems Department, California Institute of Technology, Pasadena, 1994).
6. Boahen, K. A. *Analog Integr. Circ. S.* **13**, 53–68 (1997).
7. Liu, S.-C. in *Advances in Neural Information Processing Systems* Vol. 10 (eds Jordan, M. I., Kearns, M. J. & Solla, S. A.) 712–718 (MIT Press, 1998).
8. Mahowald, M. PhD thesis, California Institute of Technology (1992).
9. Arreguit, X., van Schaik, F. A., Bauduin, F., Bidiville, M. & Raebler, E. *J. Solid-State Circ.* **31**, 1916–1921 (1996).

Faculty start-ups offer temptation to breach academic rules ...

Sir— There is a growing discomfort in the academic community about our financial and structural relations with the corporate sector, as identified in your timely Opinion article “Is the university–industrial complex out of control?” (*Nature* **409**, 119; 2001). But I believe the situation is much more complex than you have described, and the “new year’s resolutions” you offer as solutions are rather misdirected.

You describe benefits to researchers and universities, but these relate only to collaborations with large corporations, not smaller companies or faculty start-ups. In my 20-year experience, collaborations and/or contract research agreements with large companies are typically straightforward and non-controversial. Ownership issues are contractually spelt out: the university retains ownership, with various licensing rights accruing to the company. The technologies covered are specified with mutual consent. Publication delays (if requested) are specified within university policy — short delays are generally allowed to protect intellectual property rights.

There are examples that deviate from this norm. But universities do not have to make agreements that do not meet normal academic criteria. They should refuse relationships that might, for example, restrict academic freedom or improperly institutionalize a company’s influence.

What institutions need is good advice on how to forge a reasonable collaboration, and protection from its faculty members, who may be clamouring to accept finance for their research at any price. Both problems would be reduced by a manifesto in which universities would adhere to a common corporate contract, with exceptions possible and noted when necessary. Some inappropriate deals are signed because universities believe that if they don’t, some other university will.

Your editorial suggests that academics should “stand up for themselves, with the protection of university constitutions and hierarchies”. But academics are not a single entity. Corporate–university interactions are irrelevant to many, who have opposed partnerships with industry.

There are very significant differences in the relationships a company wants with an institution: commercial exploitation of some technologies depends on know-how and speed of market penetration, whereas in others it depends on a solid base of intellectual property protection. It is not clear to me what faculty members would say if they were to attempt to speak with a

single voice, other than insist on academic freedom. While laudable, that is not enough.

Where I believe university–industry relations are truly “out of control” is in our dealings with small companies, especially those founded on the basis of discoveries by faculty members where the individual and the institution both hold equity.

The benefits of corporate partnership do not typically flow from deals with small companies. These, especially faculty start-ups, rarely have valuable databases, unique facilities, market access or any assets other than the technology available through a university licence. The driving principle is not to benefit university research, but to benefit the faculty member financially. Conflicts of interest in these cases are most difficult for the university to administer.

Although virtually every university has policies to define acceptable behaviour, policies alone are not sufficient. Biased publications and undeclared conflict of interest are inappropriate actions by individual academics. And what about undeclared consulting benefits, undetected data manipulation, unregistered transfer of university intellectual property, undeclared equity interest, inappropriate use of federally supported students and postdocs on company projects, and so on?

The only “resolution” that will solve these problems is for universities to ban their academic staff from simultaneously holding equity in a company, and to severely discipline those who make corporate relationships on their own.

I am convinced that, done properly, academic–corporate collaborations benefit both partners, enhance the quality of university research and are in the public good. It is expecting too much to suppose that academics and their institutions can always act prudently and responsibly when each holds equity in the same venture. Universities should take equity because that is how they recoup the public’s investment in their research. Faculty members cannot realistically be expected to invest simultaneously in the public good and their own private financial interests.

Richard K. Koehn

*Department of Biology, University of Utah,
Salt Lake City, Utah 84112, USA*

... but Syngenta deal is a boon to Berkeley

Sir— The editorial “Is the university–industrial complex out of control?” (*Nature* **409**, 119; 2001) stated that Novartis (now Syngenta) “gains a seat in university and departmental research committees and restricts academics’ freedom to discuss the benefits of the deal”.

As graduate students of Berkeley’s plant and microbial biology department, which has made the collaborative agreement with Syngenta, we disagree. We are free to discuss the benefits of the deal, and department committees at Berkeley do not contain members from the company. There is, however, one new committee dedicated solely to awarding research funds from Syngenta, which does include company representatives.

Ironically, Berkeley was the lead player in your Opinion article, but the agreement between Berkeley and Novartis in 1998 is a model for the “new year’s resolutions” that the article listed. Information about the agreement is available at plantbio.berkeley.edu/PMB-TMRI, including limits to publication delay, restrictions on licensing, and a list of projects funded. This document illustrates that the basic research mission of the department is promoted by the collaboration.

Matthew Metz

*Department of Plant and Microbial Biology,
11 Koshland Hall, University of California,
Berkeley, California 94720, USA*

Other signatories to this letter:

J. Peter Coppinger, Carolyn Rasmussen, Jinjer Larsen,
Michael J. Axtell, Rebecca Middleton, Meredith Johnson,
Jennifer Bragg, Jennifer Johnson

‘Art’ was a load of fluff

Sir— The word ‘bollocks’ appeared in *Nature* for the first time (*Nature* **392**, 663; 1998) when Martin Kemp quoted the reaction of a lowly Leicester University student to some photographs published in Kemp’s ‘Art and Science’ series.

I was that student, and the remark ended up in Kemp’s article because, on seeing Cornelia Parker’s photographs of navel fluff and the like, I had exclaimed “What’s this bollocks doing in *Nature*?” to his daughter Dr Joanna Kemp, who was also at Leicester at the time. Unfortunately, I was at sea when the second article was published, so was unable to respond to his somewhat scathing comments.

As it is sadly likely that my association with Kemp’s article is the closest that I will ever get to a *Nature* publication, I am grateful for this opportunity to have my contribution to advancing the understanding between science and art formally acknowledged by your publication of this letter.

The fact that I am responsible for the word ‘bollocks’ first appearing in *Nature* is not only a brilliant addition to my CV but may even represent the pinnacle of my scientific career.

Magnus Johnson

*Scarborough Centre for Coastal Studies, University
of Hull, Filey Road, Scarborough YO11 3AZ, UK*



Managing foot-and-mouth

The science of controlling disease outbreaks.

**Mark Woolhouse
and Alex Donaldson**

Outbreaks of foot-and-mouth disease (FMD) have been reported from 34 different countries in the past 18 months¹, but the ongoing UK epidemic, which has already spread to France, Holland and Ireland, is the most serious of these. The introduction of the disease into the United Kingdom apparently represents a failure of, or evasion of, import controls for livestock or livestock products — although the 20 years since the last outbreak is the country's longest period of freedom from this disease during the past century. Although the origins of the current epidemic are not yet, and may never be, known², its subsequent spread raises issues about the management and control of infectious diseases in livestock, both specific to FMD and more generally.

The importance of livestock movements for rapid dissemination of FMD across the United Kingdom is already clear, with at least seven sheep and cattle markets being heavily implicated in the spread of the FMD virus³. The movement of animals is a long-recognized risk factor for the spread of infections: in the mid-eighteenth century, sheep farmers petitioned the UK parliament complaining that a serious epidemic of sheep scrapie at that time had been caused by dealers moving sheep around the country⁴. Movement restrictions are therefore an important part of disease control programmes. In the case of the current UK FMD epidemic, it remains to be determined how

much of the early spread of infection took place after 20 February, when the first outbreak was confirmed, but before the evening of 23 February, when livestock movement restrictions were imposed.

The rapid spatial spread of FMD in the early phases of the current UK epidemic is an example of 'small world' dynamics⁵, in which a population characterized by mostly local contacts is transformed into a 'well-mixed' population by the presence of remarkably few long-distance contacts (livestock markets or dealerships with big catchment areas in this case), with drastic implications for the spread of infection. Long-distance contacts evidently extend into the rest of Europe. A flourishing international livestock trade means that we are not just living in a 'global village'; we are living on a global farm. One solution to this problem would be to impose minimum periods of residency before livestock which have moved between holdings can be moved again.

Out of control

Since the imposition of movement restrictions in the United Kingdom, most transmission has been confined to local spread of the disease⁶. But a new report (see News story on page 501) concludes that the epidemic is still 'out of control' in the sense that each outbreak (defined as the occurrence of infection on a single holding) is generating an average of more than one subsequent outbreak. More formally, this observation means that the case reproduc-

tion ratio, R , is greater than one. The immediate aim of a control programme must be to reduce R to less than one.

Control of FMD relies on quarantining and 'stamping out' individual outbreaks. Usually, the policy works: FMD has been recorded in the United Kingdom in 26 of the past 59 years, but the second highest number of outbreaks during this time has already occurred this year. A key point is that the interval between a flock or herd becoming infectious and it being slaughtered must be kept to a minimum. Recent analyses of FMD epidemics indicate that reducing this interval can lead to a substantial reduction in the total size of an epidemic^{7,8}. One way of achieving this is to reduce the time between the reporting of disease and the slaughter of affected animals. This is essentially a logistic problem. During an FMD epidemic in Denmark in 1982, the interval (excluding the first outbreak) averaged 15 hours; that epidemic was contained after 22 outbreaks⁹. Early indications are that during the current UK epidemic, the interval has been approximately 2–3 days, although efforts are now being made to reduce it. It is likely that delayed slaughter times have had a significant impact on the epidemic's scale.

Another major problem is that some infected livestock (especially sheep) never develop clinical signs such as vesicular lesions and so some outbreaks may be missed altogether. Even where FMD is diagnosed, there may still have been time to transmit infection because infected animals can be infectious for some time (one or two days for

sheep) before clinical signs appear¹⁰. If holdings with undiagnosed or not yet diagnosed infections make a substantial contribution to overall transmission rates, then stamping out based on clinical signs may not by itself reduce R to less than one. There is huge interest in the development of diagnostic tests for subclinical infections of diseases with long incubation periods (for example scrapie in sheep) for exactly this reason. It is less widely appreciated that the need for early diagnosis can be just as important for diseases with short incubation periods, such as FMD. Hence a research priority is to develop a cheap, rapid pen-side diagnostic test for detecting FMD virus in the absence of clinical signs¹¹.

A stamping-out policy may include the pre-emptive culling of 'at risk' herds or flocks. The inclusiveness of the definition of 'at risk' must be correctly balanced with logistical and other concerns (especially those of animal welfare and farmers' cooperation) inherent in slaughtering large numbers of livestock. This issue is currently being hotly debated. Another crucial point is that pre-emptive culling must be undertaken quickly because of the interval between animals becoming infectious and showing clinical signs. If pre-emptive culling of 'at risk' animals is delayed by even a few days, those animals might transmit the infection to other holdings. Much the same problem applies to the vaccination of 'at risk' livestock. Because even the most potent vaccines available require three to four days to induce protection¹², if vaccinated animals can be infected and/or can transmit infection even for a limited time, then the disease will continue to spread. Vaccination, just like pre-emptive culling, should reduce R , but is unlikely to bring it down to zero.

Vaccination concerns

There are numerous other well-rehearsed problems with vaccination against FMD, including the long-term consequences of a vaccination programme for the export of livestock and livestock products¹³. There are some clear research needs here. One is for the development of a vaccine to induce sterilizing immunity in the respiratory tract of sheep and cattle that could be used without the risk of establishing carrier animals. This would counter the strongest argument against vaccination, which is that vaccinated animals exposed to virus may become carriers, leading to countries with vaccination programmes regarded by trading partners as potentially infected and therefore at risk of extended livestock trade embargoes. A much less widely appreciated need is for a vaccine that quickly stops infected animals from transmitting the infection rather than aiming to prevent disease. Most discussion of vaccine performance concerns efficacy and the duration of protection¹⁴. But in contrast



The course and scale of the current UK epidemic differs from those of its recent predecessors.

to vaccination programmes designed to prevent epidemics by providing herd immunity¹⁵, management of ongoing epidemics requires quick-acting, transmission-blocking vaccines.

For any epidemic, the value of R is critically dependent not only on the transmission biology of the infectious agent but also on the demography of the host population. The latter includes population density, patterns of movement, recruitment rates (breeding or immigration) and loss rates (death or emigration). Knowledge of host demography is essential for epidemiological research to design control programmes in advance of an epidemic. It is also necessary if we are to establish whether changes in the demography of livestock as a result of intensification of farming have contributed to the severity of the current FMD epidemic in the United Kingdom, and if they predispose the country to more serious epidemics of other livestock diseases in the future.

Some demographic information for UK livestock populations is provided by annual Ministry of Agriculture, Fisheries and Food census data⁶. But the fine detail of the spatial distribution of livestock holdings is not in the public domain, which is a major barrier to epidemiological research. Similarly, information on the age structure of livestock populations is either not available or is not in the public domain — absence of this information for UK cattle handicapped research on the epidemiology of BSE¹⁶. In the European Union, information is now collected on movements of cattle and pigs (less so for sheep), but these essential data can be equally difficult to obtain for non-government scientists.

The same demographic information is a prerequisite for attempts to forecast the course of an epidemic already under way.

Projecting epidemics is a notoriously difficult task, especially when the underlying patterns of transmission are constantly changing as control measures are introduced and refined. The accuracy of such projections is limited by three factors: (1) the quality of the data from which the projection is made; (2) the appropriateness of the underlying mathematical model of the epidemiology of the disease (and there is a wide variety of modelling approaches available); and (3) the further forward in time that the projection is made. However, projections are useful for policy-makers, not just because they provide a guide to immediate trends but because they can provide reference points against which to judge, as quickly as possible, whether changes in disease-control policies are having the desired effect.

The aim of disease control programmes is to reduce R , if not to zero then at least to below one, so that an epidemic goes into decline. But even then, the infection may not be eradicated for some time, as diminishing chains of infection result in a long 'tail' to the epidemic, as occurred in the UK 1967–68 FMD epidemic. It is unlikely that the disease could become endemic in the United Kingdom because in recent outbreaks it has been successfully eradicated from the country and from other European countries every time that it has been introduced, even when control measures have been relatively lax. That said, the course and scale of the current UK epidemic differs from those of its recent predecessors, and the elimination of the disease this time will be a formidable challenge.

Mark Woolhouse is at the Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK. Alex Donaldson is at the Institute for Animal Health, Ash Road, Pirbright, Woking, Surrey GU24 0NF, UK, which incorporates the OIE/FAO World Reference Laboratory for Foot-and-Mouth Disease. The authors have been advising the government on the management of foot-and-mouth disease.

1. <http://www.oie.int>
2. Knowles, N. J. *Vet. Rec.* **148**, 258–259 (2001).
3. Anon. *Vet. Rec.* **148**, 286–288 (2001).
4. Woolhouse, M. E. J. *et al. Trends Microbiol.* **9**, 67–70 (2001).
5. Watts, D. J. & Strogatz, S. H. *Nature* **393**, 440–442 (1998).
6. <http://www.maff.gov.uk>
7. Haydon, D. T. *et al. IMA J. Math. Appl. Med. Biol.* **14**, 1–9 (1997).
8. Howard, S. C. & Donnelly, C. A. *Res. Vet. Sci.* **69**, 189–196 (2000).
9. Westergaard, J. M. (ed.) *Report on the eradication of foot-and-mouth disease on the islands of Funen and Zealand, Denmark 1982* (Danish Veterinary Service, Copenhagen, 1982).
10. Donaldson, A. I. in *Diseases of Sheep* 3rd ed. (eds Martin, W. B. & Aitken, I. A.) 254–258 (Blackwell Science, Oxford, 1999).
11. Donaldson, A. I. & Kihm, U. *Rev. Sci. Off. Int. Epiz.* **15**, 863–873 (1996).
12. Doel, T. R. *Vaccine* **12**, 592–600 (1994).
13. Salt, J. S. in *Veterinary Vaccinology* (eds Pastoret, P.-P. *et al.*) 641–652 (Elsevier, Amsterdam, 1997).
14. Woolhouse, M. E. J. *et al. Vet. J.* **153**, 41–47 (1997).
15. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Science Publications, Oxford, 1991).
16. Anderson, R. M. *et al. Nature* **382**, 779–788 (1996).

The Bone Wars revisited

The best and the worst of palaeoanthropology in Ethiopia.

Adventures in the Bone Trade: The Race to Discover Human Ancestors in Ethiopia's Afar Depression

by Jon Kalb

Copernicus: 2000. 389 pp. £20.50, \$29

Tim D. White

In a discipline where book contracts are routinely signed before fossils are freed from their matrix, tradition dictates that palaeoanthropologists regularly inflict books about their exploits on the eager public (and wary colleagues). From Louis Leakey's autobiographies and Raymond Dart's *Adventures with the Missing Link*, to Donald Johanson and Maitland Edey's *Lucy: The Beginnings of Humankind*, the public has been treated to books that blend science, adventure, intrigue, hearsay and travelogue with the discovery and analysis of human ancestors. This book is all that and more, with a message that transcends simple entertainment or academic history, and bears directly on recent headlines about palaeoanthropological ethics.

Maurice Taieb, Yves Coppens and Donald Johanson each contributed their own 'I Love Lucy' books to this genre, based on 1970s discoveries in Ethiopia's Afar Triangle. Jon Kalb's book now extends the 'Bone Wars' trilogy and delivers what may be the last word on those early days. Kalb went to Ethiopia in 1971 and joined the explorations of this better-known trio. He left the project a year before Lucy's scientific debut in 1974, but remained in Ethiopia until 1978, where he was in a unique position to observe the development of palaeoanthropology there. How accurate is his account, and how will it play in palaeoanthropology—or in Kansas?

The book is a fascinating, well-crafted personal narrative. Kalb interweaves his palaeoanthropological experiences in the Afar with introductions to the region's geology and geography. His chronicles are skillfully contextualized against the historical backdrop of socio-political events in the Horn of Africa during the 1970s. The result is a book with something for many readers, from the historian of science to the layman. Kalb's style is engaging, and he deftly blends humour, anecdote, science, history and adventure. He showcases the best and worst of palaeoanthropology, with an important take-home message about the development of African scholarship.

Kalb, like many autobiographers, sometimes gives in to the temptation to 'improve' his tales. He believes that his problems in Ethiopia stemmed from a vast conspiracy by

the scientific establishment. After abandoning his graduate studies in geology, Kalb developed visions of becoming Africa's next famous amateur palaeoanthropologist. Richard Leakey had accomplished this in Kenya, and was by then snubbing academics by boasting that the only time he set foot in a university was to give invited lectures. But

Leakey's success in the increasingly multi-disciplinary science of palaeoanthropology was built on 40 years of family tradition and control in former British colonies. In Ethiopia, Kalb attempted to cast himself in a Richard Leakey mould. It didn't work. He fooled neither the government regulators nor the academics already doing research there.



Bony booty: Donald Johanson (left) and colleagues with the hominid fossils found in the Afar region.

Like any good hero myth, Kalb's story pits a virtuous outsider-neophyte-innocent (himself) against evil giants (Johanson, the National Science Foundation, "Berkeley types" and everyone's favourite bogeymen — peer reviewers). But whatever Kalb may have intended, it is the geologist Maurice Taieb who emerges as the true hero in this book. It was Taieb who first explored the entire Awash Valley, and found Hadar and most of the other major study areas of the Afar region. He recognized their palaeo-anthropological significance, brought them to the attention of the world, shared them with large teams of professional scientists, and changed palaeoanthropology for ever.

Kalb's unrelentingly negative characterizations of his one-time friend and partner Donald Johanson are both amusing and revealing. Readers who enjoyed Johanson's popular accounts of the Lucy discovery and its aftermath may be surprised by Kalb's warts-and-all descriptions of the same exciting days. Despite his current avowed dislike of Johanson, Kalb does admit to enjoying his company in the early years. He attributes much of this to field time spent on recreational drugs. Kalb is silent about whether his idiosyncratic ideas on Afar tectonics and lake migration originated from this source. In contrast, he makes it perfectly clear that once the Hadar hominids were found, the camaraderie with Johanson rapidly turned to one-upmanship, and worse. The reader should thus 'consider the source' when judging the contents of Kalb's book. For example, knowledgeable readers will recognize his portraits of Desmond Clark and Berhanu Abebe, among other scholars of Ethiopia, as unwarrantedly negative.

Throughout the book, Kalb calls attention to the need for continued development of personnel and infrastructure in Ethiopian palaeoanthropology. This is an important consideration too often ignored in this type of account. Kalb's most serious contention is that Western palaeoanthropologists have exploited Africa's fossil fields without giving back. In far too many instances, he is correct. The academic community too often forgets that these fossils, so critical for understanding human evolution, are ultimately Ethiopian antiquities whose study involves Ethiopian scholars.

Today, Ethiopian scholars contribute regularly and directly to advances in palaeo-anthropological knowledge and have achieved international status as professionals. A second generation is currently developing. But even now, three decades after the events chronicled by Kalb, the 'Bone Wars' rage on. Attitudes decried by Kalb remain fossilized in some palaeoanthropologists. The sad legacy of Western colonial exploitation of African palaeontological resources is not merely a historical phenomenon.

Kalb best captures only the early years of

"the race to discover human ancestors in Ethiopia's Afar Depression". As he so clearly documents, the lure of ancient human fossils can evoke the worst in modern human behaviour. Have things improved? Last year a highly qualified young Ethiopian scholar was displaced from a new hominid site he had discovered in the Afar. He had worked there for three seasons, and had even published. But without his consent or initial knowledge, and despite his protests, foreign researchers have now managed to collect and export fossils (including a hominid) from his localities. The foreign researchers claim that their acquisition of his site was entirely legal, but I consider the ethics of their approach to be questionable.

In my opinion, this kind of behaviour, all too common among modern palaeoanthropologists and government officials, spells serious trouble for human evolutionary studies. Many Western palaeoanthropologists turn a blind eye to the ethics of such cases, dismissing them as internal African bureaucratic disputes, while eagerly awaiting invitations to study the exported fossils. Kalb's will not be the final dispatch from Afar. ■

Tim D. White is at the Laboratory for Human Evolutionary Studies, Museum of Vertebrate Zoology, and Department of Integrative Biology, University of California at Berkeley, Berkeley, California 94720, USA.

Citation gold standard

The Web of Knowledge: A Festschrift in Honor of Eugene Garfield

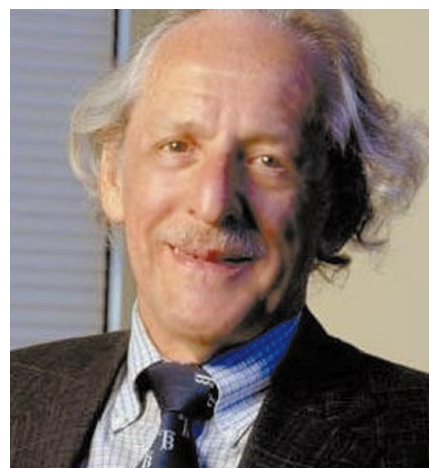
edited by Blaise Cronin & Helen Barsky Atkins

Information Today, Inc.: 2000. 544 pp. \$49.50

John Ziman

Until the 1950s, only librarians, journal subeditors and other information technicians were interested in 'the communication system' of science. True, we knew about abstract journals, where creative scrutiny of an idiosyncratic subject index might unearth something of more than passing interest. Then came Eugene Garfield, a restless postdoc wandering in the ill-charted wilderness of biomedicine. Garfield invented and then used his entrepreneurial skills to promote the *Science Citation Index (SCI)*.

Here at last was a systematic procedure for searching the literature. First, start with any paper relevant to your problem and list all the earlier sources it cites. Next, locate in the *SCI* all the more recent papers that cite the items on your list. List the most vaguely



Insightful: Garfield's *SCI* continues to prove its worth as a general catalogue of scientific activity.

relevant of these, along with the sources they cite. Iterate, until nothing more of interest turns up. Thus, surely, you will discover most of what is scientifically known, or conjectured, on the subject.

Garfield's idea was not entirely novel. In the legal profession, where precedent is holy writ, citations had been indexed semi-officially since 1821. What now made it feasible in science was the electronic computer. In 1963, with the encouragement of a few far-sighted senior scientists — especially the Nobel-prizewinning geneticist Joshua Lederberg — he published the first volume of the *SCI*. Its scope was already large enough to be of real use, and it quickly became a standard component of the scientific information system, found in every institutional library.

How well does it serve its original purpose now? Is it commercially viable as a search tool? The pharmaceutical industry must find it indispensable. I imagine that all industrial research and development organizations rely heavily on it to check for prior discovery claims relevant to their patents. But this book is suspiciously mute on the subject. Perhaps the *SCI* is already being superseded by more powerful Internet search engines.

Yet the *SCI* continues to prove its worth in other ways, especially as a general catalogue of scientific activity. Each year it lists a vast number of papers in a variety of disciplines, tagged by author(s), journals and institutions. It is much easier to retrieve, count and compare these numerically, category by category, than it ever was in conventional abstract journals. Weight these figures with 'impact factors' derived from citation rates, and you have a wonderful database for quasi-quantitative research. In other words, the *SCI* allows the systematic practice of 'scientometrics', which provides the substance of much of this book.

But there are some serious pitfalls. For example, the *SCI* doesn't cover all the

scientific literature, especially that outside the United States. Thus, as Jane Russell points out in her essay, the index seriously undervalues the work of the many excellent researchers in developing countries such as Brazil, who mainly publish and are cited in their national journals. Again, as Jonathan Cole admits, individual citations are so variable in meaning and weight that it is absurd to take notice of small numerical differences. And can Anthony van Raan really believe that the distinction between 'top' and 'not-quite-top' research performance might be represented by a numerical indicator calculated from such wonky data?

The trouble is that the supposed quantities in this assumed relationship are statistically completely abnormal. The distribution of citation rates is extremely skew. Just a few scientific papers are cited ubiquitously and endlessly: but the great majority are cited only by their proud authors.

And as Lev Landau once remarked, research performance ought to be measured logarithmically: first-class scientists are extremely rare, but they achieve 10 times as much as their numerous second-class colleagues, and so on.

In such circumstances, the arithmetical mean is not a stable statistic. Thus, it is perfectly natural mathematically for van Raan to find that impact factors seem to decline when averaged over larger and larger groups. The more elaborate indicators sometimes derived from such data are even more suspect.

Again, consider the scientometric analysis of co-citation linkages — that is, the number of times that two authors cite, or are cited by, a third one. Clearly, this is an indication of some sort of epistemic and/or social connection. The network of connections can be pruned into 'trees' (Henry Small), or scanned for 'clusters'. Experts claim (although without compelling proof) that this is a helpful way of surveying a scientific domain and getting early warning of nascent relationships. But as Tony Cawkell points out, there are many different ways of presenting these connections visually. What he fails to observe is that a scientific document has no numerical coordinates other than its date, so that even the dimensionality of this 'map' is notional. The usual choice of just two dimensions can produce geometrical contradictions. Thus, for example, Howard White's 'point-line-plane' strategy for mapping co-citational 'neighbours' of a given author is not always feasible. On the other hand, a three-dimensional map is

difficult to read, and mere humans can't grasp four-dimensional clusters. The complex multiple connectivity of the scientific literature undoubtedly says something about some of its cognitive and social features. But exploration of these is not, strictly speaking, a 'scientific' enterprise. For if there really is an underlying invariant 'structure', it is a topological abstraction that is not necessarily capable of being mapped consistently in a limited number of space-like dimensions.

Scientific citation is a thoroughly institutionalized, social practice. It functions effectively as a two-way token of interpersonal trust (Elisabeth Davenport and Blaise Cronin). In essence, it is an exchange process, where instrumental legitimacy is traded for symbolic recognition (Robert Merton). Closer attention could well be given to its changing norms. But, as Charles Oppenheim says of patent citations, "it is remarkable how much work has been undertaken based upon unproved or shaky foundations". Even after 40 years of devoted research, the foundations of Eugene Garfield's enterprising idea are still not that much more secure. ■

John Ziman is at 27 Little London Green, Oakley, Aylesbury HP18 9QL, UK.



Crescent Sun: a partial eclipse over the Hagia Sofia, or Church of the Holy Wisdom, in Istanbul, Turkey.

Darkness at noon

Glorious Eclipses: Their Past, Present and Future

by Serge Brunier & Jean-Pierre Luminet
[translated by Storm Dunlop from
Eclipses, Les Rendez-vous Célestes,
Larousse-Bordas: 1999]
Cambridge University Press: 2000. 192 pp.
£25, \$39.95

Jay M. Pasachoff

At first sight, this is the ultimate eclipse book. Oversize, well-produced photographs in a large-format book are accompanied by interesting text covering a wide variety of eclipse phenomena. The partnership of an astronomer writer/editor and a professional astronomer, albeit a cosmologist, successfully brings the beauty of total solar eclipses to the fore.

Although the writers alternate, the difference in tone is not jarring, perhaps thanks to the expert translator, Storm Dunlop. In a chapter entitled "The great cosmic clock-work", Serge Brunier, long-time editor of the French popular astronomy journal *Ciel et Espace*, discusses the many kinds of eclipses and occultations (when one astronomical body occults, or hides, another). He even includes the eagerly awaited 2004 transit of Venus — the passage of Venus across the face

of the Sun. This will be the first transit of Venus visible from Earth since 1882.

The reproduction of photographs in 25 cm × 35 cm format on heavy stock is outstanding. For example, we see Saturn's rings extending across a double page for almost half a metre in a Hubble Space Telescope image. The authors justify using this image, which shows the shadow of Saturn's satellite Titan on Saturn's clouds, by pointing out that it corresponds to an eclipse of the Sun by Titan as seen from Saturn. It's a pity, though, that Brunier calls himself an "eclipse chaser" — that common but misleading phrase seems to imply that people can outrun, or even outfly, eclipses, which has never been done.

Brunier is responsible for some of the most magnificent photographs, such as one from the 1991 total solar eclipse visible from the Mauna Kea Observatory in Hawaii, with a few telescopes in the foreground. But the image I found most arresting is not one of his: it is a wide-angle shot of the 1991 eclipse viewed through a hole in the clouds and showing the front of Reims Cathedral in the left foreground.

In his chapters on the history and cultural significance of eclipses, Jean-Pierre Luminet reproduces a remarkable set of historical images related to solar and lunar eclipses from around the world. These are no doubt related to his work on an exhibition of astronomical atlases held in Paris at the Bibliothèque Nationale in 1998.

It was good to read about myths that cast eclipses in a favourable light rather than as evil omens. Luminet also describes the history of scientific discovery through observations of solar eclipses. One example is the discovery at the 1868 total solar eclipse of the spectrum of the solar chromosphere — the layer of the Sun just above its surface that becomes briefly visible at the beginning and end of totality. The element helium was also discovered at this eclipse.

Unfortunately, though, the scientific observations of eclipses are not brought up to date. The authors describe Serge Koutchmy's wonderful work on Mauna Kea at the 1991 eclipse, but do not discuss the work of the many eclipse teams and the eclipse observations that have been made since then. Eclipse science continued in Chile and Bolivia in 1994, in India in 1995, in Aruba and across the Caribbean in 1998, and across Europe, Turkey, Syria and Iran in 1999. As a result of these omissions, the book does not give proper credit to contemporary eclipse science.

It is also a shame that the authors do not consider the current status of ground-based eclipse studies. These still have many useful observations to make that are beyond the capabilities of the solar spacecraft now aloft. And today's ground-based eclipse experimenters often work in collaboration with

instruments aboard the Solar and Heliospheric Observatory (SOHO) spacecraft and the Transition Region and Coronal Explorer (TRACE), or use space data to orient their telescopes in preparation for totality. This trend will continue in southern Africa, notably in Zambia, at the total eclipse of 21 June this year. But SOHO is mentioned in only three paragraphs and TRACE is represented by a single picture. Neither spacecraft appears in the index, which is, *ipso facto*, inadequate.

But there is much that is excellent in the book. The charts of future eclipses at the end of the book are redrawn at the highest quality. Lunar eclipses are included in the text, photos and charts, looking almost as dramatic on the page as the solar eclipses, even though they are much less so in reality. As a book of history, myth, literature, photography and expeditionary experiences, *Glorious Eclipses* is outstanding, despite its omission of the solar research carried out at recent eclipses. ■

Jay M. Pasachoff is at the Hopkins Observatory, Williams College, Williamstown, Massachusetts 01267, USA.

Feeling our way

Emotion: The Science of Sentiment

by Dylan Evans

Oxford University Press: 2001. 204 pp.

£9.99, \$15.95

Simon Baron-Cohen

This is a fun little book, no bigger than your palm, and with a bright pink cover, summarizing in a popular science format what we know about emotion. Dylan Evans is a researcher at the Darwin Centre at the London School of Economics, and his account is, predictably, about the evolution of human emotions. Highly accessible, this little gem deserves to sell well.

Typically, only about 5% of what we teach on undergraduate psychology courses is emotion, the remainder being cognition. This is despite the fact that



Let it out: all the important aspects of everyday life are governed by emotion.

all the important stuff in our everyday lives is determined by emotion. Evans' agenda is to redress this balance, and, given how student-friendly this little book is, we can hope to see emotion become more central in behavioural neuroscience research.

Evans admits that his book leaves out a fair

bit. For example, it doesn't touch on individual differences or on development. You might think that these are two big omissions, for many of the interesting questions centre on why you and I experience our emotions differently, and on the role of development in creating these individual differences. Psychotherapists will therefore find that this book contains little of relevance to their clinical work.

But the strength of the darwinian approach is to sweep aside individual differences in an effort to highlight the universals. Universals are compelling, because we can then rule out the role of culture and focus on the products of evolved biological systems. This is Evans at his best. Following Paul Ekman, the psychologist most well known for his comparative studies of emotion in Western and non-Western cultures, Evans lays out some central examples of emotion that you will find the world over, and for which one can make a convincing argument for their adaptiveness.

The obvious case of an emotion being adaptive is fear. Those of our ancestors who experienced too little fear may have stood for too long at the edge of a cliff or staring into the face of a lion, and their genes would in all likelihood have died with them. Ancestors who experienced too much fear may have been equally unlikely to reproduce if they were too afraid to even venture out of their cave. A midway level of fear seems therefore to have been adaptive and optimal. Evans makes similar arguments for other emotions, such as revenge, guilt, embarrassment, romantic love and sexual jealousy, as well as the usual ('basic') list of emotions — happiness, sadness, fear, anger, surprise and disgust.

I was somewhat surprised that there was no comprehensive list or taxonomy of emotions. It may be that Evans decided, wisely, to avoid grappling with this, given the controversies that exist in this field. There is no consensus as to how to classify emotions, because the dimensions along which they could be sorted rely on which features of a given emotion a researcher highlights. But this does not mean a taxonomy of emotions cannot be attempted and defended. In our lab, for example, we have been collecting emotion words in the adult English lexicon and have now classified close to 1,000 discrete emotions into 23 mutually exclusive categories. But Evans's excellent introductory book makes clear that we have a long way to go before we can give a full account of the richness of human emotions in terms of both their evolved and their experiential components. ■

Simon Baron-Cohen is in the Departments of Experimental Psychology and Psychiatry, Autism Research Centre, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

Unwritten knowledge

Preliterate societies depend on the wise words of the older generations.

Jared Diamond

Thanks to modern science and public health, most women *Nature* readers will live to experience menopause. About half the journal's readers of both sexes will experience their eightieth birthday, and many will reach their nineties. But severe social problems await our elderly — and also their middle-aged offspring, who must care for aged parents and young children simultaneously. The root of those problems is that modern society has no productive role for old people: they are considered useless, and both they and their caretakers know it. How did we get into this miserable situation?

It is often claimed that old age and menopause are artefacts of improved modern living conditions: that few women used to reach the age of menopause, and few individuals of either sex survived to 60. I believe instead that menopause and old age have been human hallmarks for a long time. Together with our large brain and upright posture, they played a decisive role in human evolution.

My views are based on my observations of dozens of human societies in New Guinea, many of them only recently contacted, living largely traditional lifestyles, and receiving little or no medical care. In every village, I encounter many people in their fifties (as estimated by their memories of datable events), and some in their seventies and eighties. The condition of these old people differs greatly from that of aged Westerners in three respects.

First, they remain integrated in the societies where they grew up. They live in the same huts as their children, surrounded by relatives and friends. Descriptions of retirement homes and Sunday visits provoke astonished outrage in New Guineans.

Second, younger relatives care for infirm old people in ways surprising to us. For example, many old New Guineans are virtually toothless, so they eat food that their children have chewed until soft and spat into a bowl for them.

Finally, old people are the repositories of knowledge in preliterate societies. In my field studies of New Guinea birds, I start work in a new area by gathering the oldest hunters and quizzing them. Out pour accounts of the 127 bird species (and hundreds of other animal and plant species) known to them, each given some name in the local language, with incredibly detailed information about behaviour, diet, habitat, nest and edibility or other uses.

We cannot conceive of a preliterate society's absolute dependence on old people as the equivalent of libraries.

When the hunters are stumped by my asking about some especially rare bird, they answer: "We don't know, let's ask the old man (or woman)." We go into another hut, where we find a blind and toothless old person who can describe a rare bird last seen 50 years ago.

Some of that stored information is essential to the survival of the whole village, whose members include most living relatives of the old person. The information encompasses wisdom about how to survive dangers — such as droughts, crop failures, cyclones and raids — that occur at long intervals but that could kill the whole tribe if it did not know how to react. For instance, on Rennell Island in 1976, I was quizzing local people about trees: for each tree species, what (if any) animal ate its fruit, and did people also eat it? I soon learnt that trees were divided into three categories: fruit inedible to humans, fruit normally eaten by humans, and fruit eaten only in famine times such as "after the *hungi kengi*". I had never heard of a *hungi kengi*, and my middle-aged Rennellese informants did not know what fruits to eat after it. Instead, they led me to a hut in which lay a frail old woman, who named a dozen fruits to eat during that time.

It turned out that the *hungi kengi* was the most destructive cyclone in Rennellese history, datable to around 1910. It had destroyed gardens, forcing people to rely on 'famine foods' that were normally not eaten but that old survivors of the previous cyclone remembered as safe and nutritious. In 1976 that old woman must have been in her early eighties, because she said that she had been a child not quite of marriageable age when the *hungi kengi* struck. Upon her knowledge would hang the survival of villagers bearing her genes, when another cyclone should strike.

For us moderns, all information essential to survival is transmitted in writing. We cannot conceive of a preliterate society's absolute dependence on old people as the equivalent of libraries. We forget that



Wisdom of age: the knowledge of elders is passed by word of mouth in preliterate societies.

CORBIS

all human societies were preliterate throughout 99.9% of human evolution, from our evolutionary divergence from apes five million years ago until the first writing around 3400 BC. As a species, we are unique in the degree to which our survival depends on detailed knowledge transmitted culturally through language rather than genetically. We are also unusual even among primates in our large and persistent social groups, complex social relationships, and long lives.

I suspect that the value of old people to their relatives' survival arose with language and was the selective factor behind the evolutionary slowdown of human senescence. It also drove the evolution of human female menopause, because a woman's risk of death in childbirth and lactation was traditionally high and increased with age, whereas men never die from childbirth and lactation. Similar considerations may have driven the evolution of menopause in killer whales and pilot whales, which also live in lifetime groups with complex relationships and culturally transmitted survival techniques. While watching individually recognized killer whales recently, I saw an 85-year-old, long post-menopausal, female still accompanied by her 55-year-old son. How can we now restore social value to old people and cope with the tragedy of growing old in our literate societies? ■

Jared Diamond is in the Physiology Department, University of California Medical School, Los Angeles, California 90095-1751, USA.

The good and the bad

Sidney Toby

A factor that may prejudice my viewpoint is my age, some 2.2 gigaseconds. The most striking changes in my lifetime have been due to the digital computer. I have seen a vast panorama of calculational technology, starting with tables of four-figure logarithms (leaflets), five figures (booklets) and seven figures (tomes). Later came linear and circular slide rules, clanky electro-mechanical calculators, the electronic desk calculator, the mainframe computer, the mid-sized computer, the hand calculator, the programmable hand calculator, finally desktop machines: fast, powerful and cheap. Molecular processors will achieve yet more.

As an educator, I can set higher-quality examinations now because students have hand calculators. As a scientist, I no longer have to spend hours doing repetitive calculations on laboratory data and plotting the results by hand. I have a large stack of linear, semilog and log-log graph paper that I cannot bring myself to throw out, even though I shall never use it again. Data processing is quicker, more accurate and more fun. At home, our television set is little used but our two computers are used constantly for data processing, writing, sending e-mail, filing taxes and, yes, scanning the web or playing games. In view of these great advantages, it seems somewhat churlish to voice criticism of computers. But there are negative aspects.

The advent of frequency, rather than amplitude, modulation improved the quality of broadcast communication, which was much further improved by satellites and signal digitization, allowing a vast increase in distance learning. Although cheap and efficient, this can be a pedagogical disaster if

economics are the sole driving force. On the other hand, recognition that most students learn best by interaction has led to effective instructor/student and student/student e-mail and newsgroups.

Kelvin's dictum, that if we can't put a number to something we don't really understand it, is true for most of science, but there are limitations. The great discoveries have been in developing a theory to fit (or, more spectacularly, predict) experimental results, best exemplified by the prediction of the existence of Neptune, Darwin's synthesis, Mendeleev's description of undiscovered elements, Einstein's gravitational bending of light and Bohr's explanation of Balmer's solar spectral lines. Computers would only have speeded up the calculations. Deciphering the genetic code, on the other hand, could not have happened without the computer: it is simply awesome bookkeeping.

Large-scale syntheses of new pharmaceutical products are now routine. Typically, an array of 96 samples — it can be as high as 15,000 — differing in one or two characteristics is automatically reacted and sampled using gas-chromatographic and mass-spectrometric analysis to find a substance with a particular property. Digitization and large-scale data storage are essential because of the enormous amount of information generated. But we are awash in data: almost every day we read of some new potential cure for a disease when all that has actually happened is that it has become easier to analyse data with statistics software and find weak correlations.

Computer simulation of reacting systems has enabled great advances in fields where change can be represented analytically. In chemistry, for example, the approximations that hobbled chemical kinetics for most of

Digitization

"Computer simulation is seductive, but it does not banish empiricism. An intellect is still required to see what is truly significant."

the past century have been abandoned. The concept of reaction order is valid for elementary steps in a complex mechanism but is meaningless for most overall reactions of the real world. Similarly, steady-state and quasi-equilibrium hypotheses are unnecessary. Computer simulations of chemical systems model the behaviour of the statistical averages of ensembles with populations of perhaps 10^{20} particles. Biological populations of cells or animals are typically more than ten orders of magnitude smaller, but the techniques are still valid. It is now relatively easy to simulate, say, complex cell growth with a sufficiently convoluted rate law. Computer simulation is seductive, but it does not banish empiricism, and when change is represented by a very complex rate law, a healthy application of Occam's razor may be needed.

The predator-prey oscillations observed by Hudson Bay Company trappers a century ago require partial second-order differential equations for their simulation, which are then easily solved on the computer. The simulation of traffic flow has recently made important strides. Among gas molecules, the frequency of collisions is proportional to the square of the particle density, but this may not be true for (far smaller) populations of cars. Monte Carlo techniques show that noise level, and thereby uncertainty, increases as the population decreases. This sets a limit to the validity of computer simulation.

The digital computer has allowed unprecedented levels of accuracy in data transmission and capacity in data storage. Let us revel in the new technology, but not uncritically embrace the results. An intellect is still required to separate what is truly significant from background noise; digitization does not guarantee understanding. We could store every musical theme that has ever been written in a few gigabytes, but the computer would not be able to generate a single beautiful melody until we can define one. ■

Sidney Toby is in the Department of Chemistry, Rutgers, the State University, Piscataway, New Jersey 08854, USA.

FURTHER READING

Maddox, J. *What Remains to be Discovered* (The Free Press, New York, 1998).

Penrose, R. *The Emperor's New Mind* (Oxford Univ. Press, 1991).

Toby, S. & Toby, F. S. The Simulation of Dynamic Systems. *J. Chem. Educ.* **76**, 1584 (1999).

Magnetic explosions in space

James F. Drake

The magnetic field that surrounds the Earth is rarely quiet. An explanation for the explosive nature of magnetic storms is gathering support from satellite data.

Violent magnetic storms at the fringes of the Earth's atmosphere are the consequence of explosions driven by an unusual source of energy — the magnetic field produced by electric currents streaming through the tenuous ionized gases (plasma), which form the Earth's magnetosphere. These storms can last from half an hour to several days, producing bright auroral displays and sometimes disrupting satellite communications (Fig. 1). Understanding how magnetic energy can be released so quickly in these storms — and in similar explosions in the Sun's atmosphere called solar flares — has perplexed scientists for nearly four decades. In the past few years a new theory has been developed to explain the explosive release of this magnetic energy. Now, on page 557 of this issue, Deng and Matsumoto¹ provide evidence for the smoking gun of the new theory — the 'whistler waves' that play a central role in the energy release².

Magnetic energy is generated throughout the Universe by the swirling motion of ion-

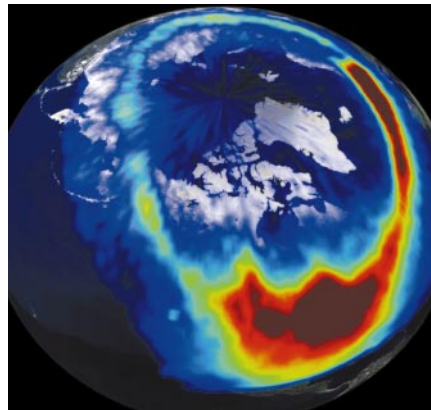


Figure 1 Magnetic storm in the Northern Hemisphere imaged by the Visible Imaging System on the Polar spacecraft in July 2000. The aurorae are represented in false colours with red indicating areas where activity is highest.

ized gases in a process known as a dynamo — the twisting and folding of the magnetic field by the plasma motion amplifies the magnetic

field. Similar dynamo processes are active in the core of most planets, including the Earth. Understanding the reverse process — the conversion of magnetic energy into the kinetic energy of the plasma flow — is crucial to both fundamental and practical science issues. For example, why are the atmospheres (coronae) of stars hotter than their surfaces? And why do plasmas in fusion experiments sometimes blow apart?

A major mystery is why the release of magnetic energy occurs so quickly. It was recognized long ago that solar flares are driven by the magnetic energy in the Sun's corona — the only available energy source that is large enough. But it takes electric currents in the corona around 10,000 years to release their energy, which is hopelessly slow compared with a flare growth time of 30 minutes or less. Similar processes with fast energy release have been documented in laboratory fusion experiments.

It is widely believed^{3–6} that the critical explosive events occur in regions where the magnetic field reverses, resulting in a magnetic 'x-line' configuration — Fig. 2 shows two examples of x-lines. Magnetic field lines act like rubber bands — when they are strongly bent they release their tension by springing outwards (Fig. 2a; black arrows). Essentially all of the magnetic energy is released from the field as this happens. Moreover, the outward-flowing magnetic field and plasma, which are tied together, cause the plasma pressure near the magnetic x-line to drop, pulling the plasma above and below towards the x-line (white arrows). So the whole process is self-driven and explosive. But it differs from a conventional explosion in that the energy is not released equally in all directions. Instead, the plasma flows in from one direction and flows out in another.

This theory, known as magnetic reconnection, is strongly supported by observation. Satellites in the Earth's magnetosphere have measured the outflow velocity of plasma on the reconnected field lines in Fig. 2a. The measurements agree with the expected Alfvén speed, c_A , which is the velocity resulting from the conversion of magnetic energy into kinetic energy. Simultaneous measurements from two satellites also confirm the geometry of the outflow — in two opposite directions⁷ (Fig. 2a).

A fundamental problem with magnetic reconnection, however, is that the theoretic-

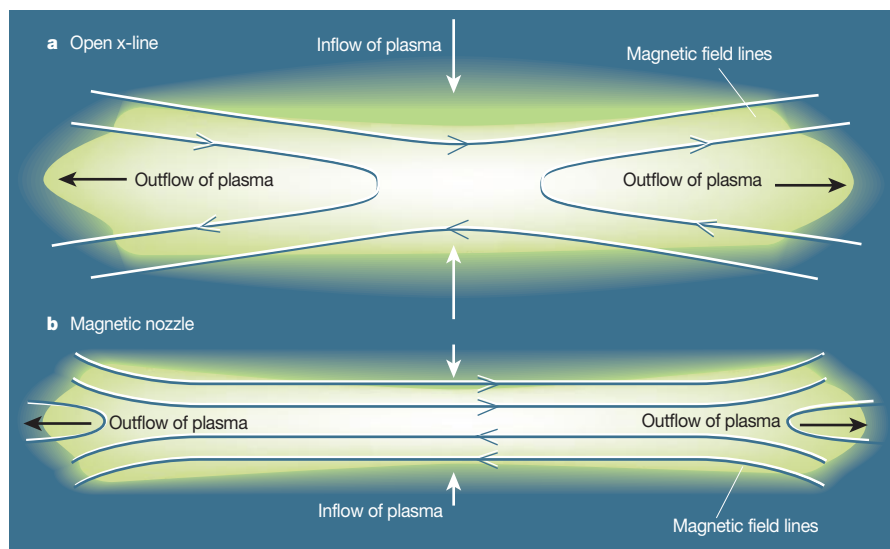


Figure 2 What happens to magnetic field lines and plasma flows during magnetic explosions. a, Open x-line configuration. In regions where the magnetic field reverses, the magnetic field lines can break and cross-link (magnetic reconnection). In this situation, the highly bent field lines appear to behave like a slingshot, first stretching then rebounding, and releasing their stored-up energy. b, Magnetic nozzle configuration. In the simple magnetized fluid picture of reconnection, the plasma forms a magnetic nozzle, and the rate of energy release is limited by the speed at which the plasma can flow out of the nozzle. This theory falls short of explaining the short timescales associated with explosive magnetic storms. Deng and Matsumoto¹ now provide evidence for a class of plasma waves called whistler waves that can accelerate electrons to much higher velocities. The presence of whistler waves means that the size of the nozzle no longer throttles the total plasma flow — the outflow velocity can increase, so the inflow rate remains high.

cal details don't add up. Instead of the open x-line configuration of Fig. 2a, the theory of simple magnetized fluids predicts a macroscopic magnetic 'nozzle' as shown in Fig. 2b. In this picture, the rate of energy release is limited by the speed at which plasma, and the magnetic field associated with it, can flow into the x-line region. Like water in a hose, the flow of plasma into the nozzle is limited by how fast the plasma flows out of the nozzle. But the maximum velocity that the outflowing plasma can reach is the Alfvén velocity. Because the outflow velocity is fixed, if the length of the nozzle in Fig. 2b is doubled, the inflow velocity is halved. So because the magnetic nozzle is very long and narrow, the inflow of plasma is severely limited and the rate of magnetic energy release is very slow. Sweet⁴ and Parker⁵ first proposed the theory behind this magnetic nozzle in the 1950s, and modern computer simulations confirm the basic finding. The problem is that the slow rates of energy release predicted by this theory are inconsistent with the explosive events seen in nature and in the laboratory.

This dilemma has been resolved in recent years as theorists re-examined the assumptions behind the simple magnetized fluid, or 'magnetohydrodynamic', picture of reconnection. The breaking of the magnetic field lines actually takes place in a narrow region around the x-line of Fig. 2a. In this narrow region the magnetohydrodynamic description fails: instead of the electrons and protons moving together, as predicted by the magnetohydrodynamic model, the two species move independently. As a result, a new class of plasma waves called whistlers is generated, which can accelerate electrons to much higher velocities in a localized region near the x-line. As the name implies, these waves were first heard by radio operators as a falling tone — the wave velocity varies inversely with the wavelength, so short-wavelength, high-frequency waves reach the operators first. The acceleration of outflowing electrons to very high velocities by the whistler wave bypasses the bottleneck created by the magnetic nozzle in Fig. 2b and restores the open x-line of Fig. 2a, with its explosive rates of magnetic energy release.

The characteristic signature of a whistler wave is the generation of magnetic field components in and out of the plane shown in Fig. 2a. Confirming the existence of these field components around the x-line is a key test of the whistler theory. In earlier laboratory experiments on magnetic reconnection, Stenzel and Gekelmann⁸ measured the out-of-plane magnetic fields. But Deng and Matsumoto¹ report the first measurements of these fields in a natural environment — from data taken by the Geotail satellite at the magnetopause, the boundary between the incoming solar wind and the Earth's magnetosphere. The data are tantalizing because the character of the magnetic field components seems to match exactly the theoretical

predictions. More detailed measurements, especially of electron and ion flows, are required to make a compelling case that the theory is valid. The proposed NASA Magnetospheric Multiscale Mission is being designed to tackle these issues by using a cluster of satellites to obtain a complete picture of magnetic reconnection.

The emerging theory of magnetic reconnection seems to offer an explanation for the fast release of magnetic energy, but leaves many mysteries unsolved. Why do the magnetic fields remain apparently quiescent for long periods of time and then suddenly explode for no apparent reason? This behaviour is seen in the solar corona, the magnetosphere and laboratory plasmas. The spark that ignites the release

process is still poorly understood. So, despite the large steps taken to unravel one of the biggest puzzles in plasma physics, the drama in our skies will demand further investigations.

James F. Drake is in the Department of Physics and the Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA.

e-mail: drake@plasma.umd.edu

1. Deng, X. H. & Matsumoto, H. *Nature* **410**, 557–560 (2001).
2. Birn, J. et al. *J. Geophys. Res.* **106**, 3715–3719 (2001).
3. Dungey, J. W. *Phil. Mag.* **44**, 725 (1953).
4. Sweet, P. A. *Nuovo Cim.* (suppl.) **8**, 188 (1958).
5. Parker, E. N. *J. Geophys. Res.* **62**, 509 (1957).
6. Petechek, H. in *AAS-NASA Symposium on the Physics of Solar Flares* (ed. Ness, W.) 425–439 (NASA, Washington DC, 1964).
7. Phan, T. et al. *Nature* **404**, 848–850 (2000).
8. Stenzel, R. L. & Gekelmann, W. *J. Geophys. Res.* **86**, 649–658 (1981).

Palaeoanthropology

Our newest oldest ancestor?

Leslie C. Aiello and Mark Collard

These are exciting times in the study of human origins. But excitement needs to be tempered with caution in assessing the claim of a six-million-year-old direct ancestor of modern humans.

Last month a team of French and Kenyan researchers, led by Brigitte Senut and Martin Pickford, published their discovery of 12 fossils, including fragmentary thigh and arm bones as well as several teeth, that they claim belong to a previously unrecognized genus and species of human ancestor^{1,2}. Originally dubbed 'Millennium man' or 'Millennium ancestor', these fossils come

from the Lukeino Formation of the Tugen Hills in Kenya. Together with an enigmatic lower left molar discovered in the 1970s, they have now been given the formal name *Orrorin tugenensis*¹.

The announcement of *Orrorin* has caused a considerable stir. This is partly because the fossiliferous deposits of the Tugen Hills are being prospected by two competing groups

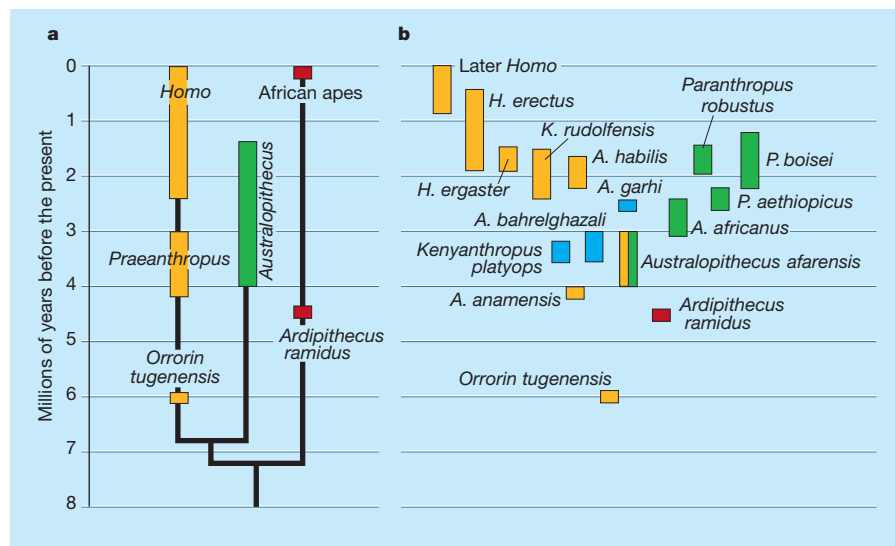


Figure 1 Hominins old and new. a, The evolutionary relationships proposed by Senut and colleagues^{1,9}. Here, the newly described *Orrorin tugenensis* is on the direct line leading to modern humans (yellow), and *Ardipithecus ramidus* is an ancestor of the African apes (red). *Kenyanthropus rudolfensis* (previously *Homo rudolfensis*) is not shown here but is included in the *Orrorin/Praeanthropus/Homo* lineage⁹. b, The known species diversity in the hominin fossil record. The species *Kenyanthropus platyops*, *Australopithecus garhi* and *A. bahrelghazali* (blue) are either too recently described or too fragmentary to have been considered by Senut and colleagues.

of researchers^{3,4}, but primarily because *Orrorin* is claimed to be about 6 million years old. This makes it 1.5 million years older than *Ardipithecus ramidus*⁵, the oldest previously recognized candidate for the earliest hominin. (Hominins include modern humans and fossil species more closely related to them than any other living species.) Significantly, *Orrorin*'s age falls within the molecularly determined range of the last common ancestor between humans and the African apes (8–5 million years ago). The authors also argue that *Orrorin* is on the direct line leading to modern humans, whereas most of the members of the genus *Australopithecus* are not (Fig. 1a). Furthermore, they reposition *Ardipithecus* as an ancestor of the African apes, rather than as the first known human ancestor.

The great age of *Orrorin* does not seem to be in serious question. The geology of the Lukeino Formation is well known²; the volcanic tuffs in this formation have been securely dated at 6.2–5.6 million years old by radiometric techniques⁶; and there is little doubt that the specimens come from the Lukeino Formation sediments². It is difficult, though, to have the same confidence in Senut and colleague's conclusions about human evolutionary history. They adopt a simple two-branch evolutionary tree for the hominins (Fig. 1a). One branch leads from *Orrorin* to *Homo* through the novel intermediary genus *Praeanthropus*; the other leads to *Australopithecus* and extinction. This simple phylogeny contrasts starkly with mainstream ideas about human evolution, and glosses over many areas of controversy and uncertainty.

The picture is further complicated by last week's announcement of yet another new hominin genus and species, the 3.5–3.2-million-year-old *Kenyanthropus platyops* from West Turkana in Kenya^{7,8}. At least 13 known hominin species from Africa existed before *Homo erectus*, and this period of our evolutionary history now looks more like a tangled bush than a simple tree (Fig. 1b). There are only a few points of consensus among most palaeoanthropologists. One is that there are four hominin genera (*Ardipithecus*, *Australopithecus*, *Paranthropus* and *Homo*), with the new *Kenyanthropus* making five. Another is that the big-toothed and massive-jawed genus *Paranthropus* (*P. aethiopicus*, *P. robustus* and *P. boisei*) represents a dead-end branch of the bush.

To understand Senut and colleagues' interpretation of *Orrorin* it is necessary to appreciate their reasons for creating an additional hominin genus, *Praeanthropus*. Senut has long believed that the skeleton, and not the skull and teeth, is the best guide to hominin evolutionary relationships⁹. She argues that the skeletal evidence suggests a very old division in hominin locomotor ability. One lineage, characterized by climbing and bent-legged bipedal walking, led to most

of the members of the genus *Australopithecus* (including *Paranthropus*); the other lineage, comprising straight-legged walkers, led from other members of the genus *Australopithecus* through *Praeanthropus* and *Homo rudolfensis* (now *Kenyanthropus rudolfensis*⁷) to *Homo sapiens*. The genus *Praeanthropus* represents those members of the genus *Australopithecus* that Senut interprets as having a skeleton suggesting more modern walking (*A. anamensis*, and some fossils normally included in the species *A. afarensis*)⁹. She also suggests that this phylogeny is supported by evidence from the teeth and jaws.

Most palaeoanthropologists do not recognize a major dichotomy in hominin locomotor ability before the evolution of *Homo ergaster*, around 1.9 million years ago, and recent analyses of the *A. anamensis* skeleton suggest that it was much like that of other members of the genus *Australopithecus*^{10,11}. Senut's claim for more modern walking for *Orrorin*, linking it with *Praeanthropus* and *Homo*, is based on detailed aspects of the anatomy of the upper part of the thigh-bone that are open to alternative explanations. For example, she and her colleagues argue that the head of the thigh-bone is very large and human-like in relation to the size of the neck of the bone. This is true, but it is also the case that *Orrorin* is no more similar to *Praeanthropus* in this feature than it is to *Australopithecus*, which also falls within the human range of human variation.

Senut and colleagues claim¹ that *Orrorin*'s relatively small, thick-enamelled molars support their interpretation that *Orrorin* is a direct ancestor of modern humans to the exclusion of *Ardipithecus* and most members of the genus *Australopithecus*. But tooth size and enamel thickness correlate with diet¹²; in the absence of other compelling evidence to link *Ardipithecus* with the African apes, or *Orrorin* with humans, it is premature to make such bold claims.

The age of *Orrorin* undoubtedly makes it a highly important addition to the debate about human origins. But we are a long way from a consensus on its role in human evolution. There are many alternative hypotheses that are equally defensible, including some in which *Orrorin* is not a hominin. *Kenyanthropus* and other new, and as yet little known, hominins such as *Australopithecus garhi*¹³ introduce further uncertainty. It also appears that cranial and dental anatomy does not necessarily mirror molecularly determined phylogenies in modern primates¹⁴, which casts considerable uncertainty on anatomically based evolutionary trees. For now, at least, it is probably best to avoid naming ancestors, and maintain a simple division: that between hominins of archaic aspect (*Orrorin*, *Ardipithecus*, *Australopithecus* — including *Paranthropus* — and *Kenyanthropus*) and hominins of modern aspect (*Homo sapiens* and the remaining species of *Homo*)^{15,16}.

Leslie C. Aiello and Mark Collard are in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK. e-mail: l.aiello@ucl.ac.uk

1. Senut, B. et al. *C. R. Acad. Sci. Paris* **332**, 137–144 (2001).
2. Pickford, M. & Senut, B. *C. R. Acad. Sci. Paris* **332**, 145–152 (2001).
3. Balter, M. *Science* **291**, 1460–1461 (2001).
4. Butler, D. *Nature* **410**, 508–509 (2001).
5. White, T. D., Suwa, G. & Asfaw, B. *Nature* **371**, 306–312 (1994).
6. Hill, A. in *Late Cenozoic Environments and Hominin Evolution: A Tribute to Bill Bishop* (eds Andrews, P. & Banham, P.) 85–97 (Geological Society, London, 1999).
7. Leakey, M. et al. *Nature* **410**, 433–440 (2001).
8. Lieberman, D. E. *Nature* **410**, 419–420 (2001).
9. Senut, B. *S. Afr. J. Sci.* **92**, 165–166 (1996).
10. Ward, C., Leakey, M. & Walker, A. *Evol. Anthropol.* **7**, 197–205 (1999).
11. Lague, M. R. & Jungers, W. L. *Am. J. Phys. Anthropol.* **101**, 401–427 (1996).
12. Teaford, M. F. & Ungar, P. S. *Proc. Natl Acad. Sci. USA* **97**, 13506–13511 (2000).
13. Asfaw, B. *Science* **284**, 629–635 (1999).
14. Collard, M. & Wood, B. A. *Proc. Natl Acad. Sci. USA* **97**, 5003–5006 (2000).
15. Wood, B. A. & Collard, M. *Science* **284**, 65–71 (1999).
16. Pilbeam, D. *Molec. Phylogenet. Evol.* **5**, 155–168 (1996).

Neurobiology

Cannabinoids act backwards

MacDonald J. Christie and Christopher W. Vaughan

Cannabis is useful for treating many ailments, but has unwanted side effects. Drugs that control signalling by cannabinoids found naturally in the body might be more useful.

Preparations from the plant *Cannabis sativa* have been used since antiquity, not only for their intoxicating effects, but also to treat a number of ailments^{1,2}. The main active component of these preparations, Δ^9 -tetrahydrocannabinol, produces most of its effects on the central nervous system by interacting with specific cannabinoid receptors on nerve cells. Under normal circumstances, these receptors are thought to

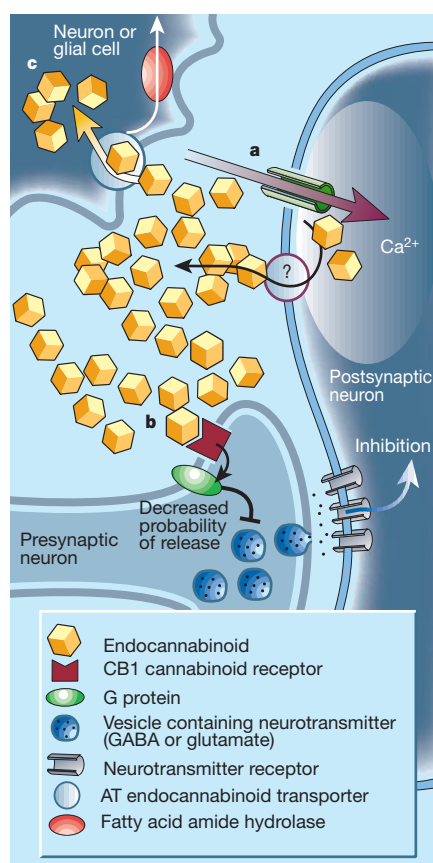
be one element of a neurotransmitter system that controls neuronal excitability. Other components of this putative signalling system include cannabinoids that are found naturally in the body, as well as cellular mechanisms by which these 'endocannabinoids' are synthesized, transported and metabolized³. But it has not been clear how important this system really is, because of a lack of direct evidence for the synthesis,

release and effects of endocannabinoids at the junctions between nerve cells (synapses) under natural conditions.

Writing on page 588 of this issue⁴, Wilson and Nicoll provide an answer. At synapses, one neuron (the presynaptic one) releases neurotransmitter molecules that diffuse to another (postsynaptic) neuron, either inhibiting or stimulating it. Wilson and Nicoll show that endocannabinoids can be formed in single postsynaptic neurons in response to physiologically relevant stimuli, and that the endocannabinoids diffuse back from the stimulated cell to act on receptors on a presynaptic neuron. This has the effect of decreasing inhibitory inputs to the postsynaptic cell. The study also highlights the importance of endocannabinoid transporters in controlling this process. Meanwhile, Ohno-Shosaku and colleagues⁵ and Kreitzer and Regehr⁶, writing in *Neuron*, show that such 'retrograde' endocannabinoid signalling is a general phenomenon, occurring at certain excitatory and inhibitory synapses in the hippocampal region of the brain, and at some excitatory synapses in the cerebellum.

It was already known that strong depolarization (excitation) of so-called pyramidal neurons in the hippocampus suppresses inhibitory inputs to pyramidal neurons from presynaptic cells. This phenomenon, known as depolarization-induced suppression of inhibition (DSI)⁷, requires an influx of calcium ions into the postsynaptic neuron. It was also known that the synthesis of endocannabinoids in neurons is stimulated by an increase in calcium concentration, and that the most common type of cannabinoid receptors, called CB1 receptors, are expressed on inhibitory presynaptic nerve terminals that form synapses with pyramidal neurons.

The three groups of authors⁴⁻⁶ have integrated these findings and implicated an endocannabinoid acting on the CB1 receptor as the retrograde messenger that is released from a depolarized postsynaptic neuron to produce DSI (Fig. 1). They established this by using synthetic molecules that antagonize the CB1 receptor or mimic the effects of endocannabinoids. They also used a synthetic molecule that binds to and activates the CB1 receptor and so prevents further activation of the receptor by endocannabinoids. Furthermore, Wilson and Nicoll⁴ and Ohno-Shosaku *et al.*⁵ find that, in response to depolarization or induction of action potentials in single hippocampal pyramidal neurons, increases in calcium concentration occur that are sufficient to trigger the production of endocannabinoids, which act as retrograde messengers. These studies have incidentally resolved another controversy by showing that another candidate neurotransmitter, glutamate, is not involved in DSI.



The endocannabinoids that produce these retrograde signals have not yet been identified. But Wilson and Nicoll have a possible answer. They show that a transporter protein called AT that is found on nearby neurons or on glial (non-neuronal) cells, and which mops up the endocannabinoids anandamide and 2-arachidonylglycerol from the synapse, is likely to be important in retrograde signalling. The authors found that AM404 — a drug that inhibits this transporter — both mimics and partially prevents DSI. How could this come about?

Presumably, by inhibiting the transporter, the drug initially prevents the endocannabinoids from being removed from the synapse, and so enables them to activate CB1 receptors. The lingering endocannabinoids acting on the CB1 receptor might then prevent further activation of the receptor by endocannabinoids produced by DSI. The significance of this endocannabinoid transporter was previously unclear, because the known endocannabinoids are lipophilic, and so were thought to diffuse freely across neuronal and glial membranes. But Wilson and Nicoll now show that the transporter is physiologically important for ending endocannabinoid signals.

These results also suggest that anandamide and 2-arachidonylglycerol are the key endocannabinoids in retrograde signalling, but this is by no means certain. Experimentally manipulating enzymes that

Figure 1 Cannabinoids found naturally in the body (endocannabinoids) act as retrograde messengers within the brain. The model shown here is based on three new papers⁴⁻⁶.

a, Excitation of a neuron causes its depolarization and an influx of calcium ions. This stimulates various phospholipases, which start the synthesis of endocannabinoids, such as anandamide and 2-arachidonylglycerol. These are released from the neuron by an unknown mechanism (?). b, Endocannabinoids freely diffuse away to bind to CB1-type cannabinoid receptors on the presynaptic terminals of neurons that form synapses with the stimulated neuron. This reduces the probability of inhibitory neurotransmitters being released. The effects of the endocannabinoids are mimicked by the active component of cannabis, Δ^9 -tetrahydrocannabinol, as well as by synthetic agonists, and are blocked by antagonists of the CB1 receptor (not shown). c, Endocannabinoids are taken up into neuronal and glial cells by a transporter and then broken down, possibly by the membrane-bound fatty acid amide hydrolase. So the amount of endocannabinoid available to regulate presynaptic CB1 receptors is regulated by uptake and degradation.

might be involved in the synthesis and breakdown of endocannabinoids — for example, by using inhibitors of the enzyme fatty acid amide hydrolase — may help to narrow the range of candidate endocannabinoids. Moreover, the mechanism by which endocannabinoids are exported from a stimulated neuron needs to be determined. It is also unknown whether endocannabinoids can be generated physiologically using stimuli other than depolarization-induced calcium entry. For example, perhaps their production might be activated by other neurotransmitter receptors that are coupled to signalling molecules called G_q and G_{11} proteins, and hence to phospholipase enzymes.

Wilson and Nicoll⁴ have also used the parallel alignment of hippocampal pyramidal neurons and their primary dendrites (extensions) to estimate the range of diffusion of the endocannabinoids released from a single neuron in a brain slice. They found that the retrograde signal affects synapses within a radius of about 20 micrometres from the stimulated neuron. But the effects of inhibiting endocannabinoid transport or metabolism on this range have yet to be studied. Such inhibitors might increase the range or intensity of retrograde endocannabinoid signalling at this or other synapses. If, as suggested by Wilson and Nicoll, DSI enhances learning in the hippocampus, then these inhibitors might intensify the process. By contrast, drugs such as Δ^9 -tetrahydrocannabinol, which act

directly on all cannabinoid receptors in the hippocampus, have disruptive effects.

Natural and synthetic cannabinoids relieve nausea and vomiting, and stimulate appetite. They also have a range of effects that suggest they might be useful in treating pain, migraine, muscle spasms associated with multiple sclerosis, and glaucoma^{1,2}. But cannabinoids also have many unwanted side effects. Our ever-increasing understanding of endocannabinoid signalling has raised hopes that useful drugs without these side effects could be developed^{1,2}.

Unfortunately, the CB1 receptor is widely distributed throughout the brain and accounts for almost all of the effects of cannabis on memory, cognition, coordination, mood, pain sensation, appetite and sleep. (The only other cannabinoid receptor identified so far, the CB2 receptor, is expressed largely in immune tissues.) So it is unlikely that synthetic drugs that activate the

CB1 receptor could sidestep all the unwanted psychotropic effects of cannabis. But drugs that inhibit endocannabinoid transporters would work only to enhance the actions of naturally released endocannabinoids, so might be much more useful.

MacDonald J. Christie and Christopher W. Vaughan are in the Department of Pharmacology and The Medical Foundation, The University of Sydney, New South Wales 2006, Australia.
e-mail: macc@pharmacol.usyd.edu.au

1. House of Lords Select Committee on Science and Technology *Cannabis: The Scientific and Medical Evidence* (The Stationery Office, London, 1998).
2. National Institutes of Health *Report on the Medical Uses of Marijuana*
www.nih.gov/news/medmarijuana/MedicalMarijuana.htm (1997).
3. Piomelli, D., Giuffrida, A., Calignano, A. & Fonseca, R. *Trends Pharmacol. Sci.* **21**, 218–224 (2000).
4. Wilson, R. I. & Nicoll, R. A. *Nature* **410**, 588–592 (2001).
5. Ohno-Shosaku, T., Maejima, T. & Kano, M. *Neuron* **29**, 729–738 (2001).
6. Kreitzer, A. C. & Regehr, W. G. *Neuron* **29**, 717–727 (2001).
7. Piltner, T. A. & Alger, B. E. *J. Neurosci.* **12**, 4122–4132 (1992).

Plant biology

Night moves of pregnant moths

Clarence A. Ryan

Tobacco plants attacked by caterpillars release different blends of volatile compounds by day and night. Those released at night tell nocturnal moths not to approach — a signal that benefits both plants and moths.

Volatile chemicals are the language of plants. Through the smell of fresh blossoms, good coffee or a fine wine, their message to humans can be attractive. But plants do not expend valuable energy making these chemicals simply to please humans, and most volatiles have more serious functions. Some, for instance, are important in communicating information to particular insects that is crucial to the survival of the plants, and often the insects as well.

On page 577 of this issue¹, De Moraes and colleagues describe a previously unknown chemical conversation between plants and herbivorous insects (Fig. 1). At night, tobacco plants that are being attacked by caterpillars emit a specific blend of volatile chemicals. Nocturnal moths interpret these chemicals as a signal that they will not be welcome to lay their eggs there. But it isn't just the plant that benefits from these night-time emissions. As the plant is making nasty chemicals to ward off the caterpillars, and may be summoning help from predatory insects, it is advantageous for the moths to keep away.

It is well known that, when being grazed on by herbivorous insect larvae, plants release volatile chemicals such as terpenoids to attract carnivorous insects that will prey on the larvae. For instance, lima-bean plants infested by spider mites release volatile chemicals that attract predatory mites,

which in turn feed on the spider mites². The chemicals produced are specific to the type of damage: neither leaves from uninfected plants nor leaves damaged mechanically in the absence of mites release these particular chemicals, so they do not attract the predatory mites. Similarly, corn seedlings and cotton plants damaged by caterpillars release volatile signals that attract parasitic wasps, which lay their eggs in the caterpillars³. When the eggs hatch, the larvae dine on their hosts, eventually killing them. These clever ruses contribute to the survival of both the plants and the predators, which do not eat leafy plants. More than 15 plant species, 10 herbivorous insect species and over 10 predatory insect species are known to be involved in such interactions⁴.

The process of attracting predatory insects involves the interaction of specific blends of plant volatiles with highly sensitive receptor molecules of the predators. And, taking a step back, the grazing herbivores themselves release chemicals in their saliva that lead to the synthesis and release of the plant volatiles.

For example, a chemical signal from the saliva of beet armyworm caterpillars has been isolated and characterized as *N*-(17-hydroxylinolenoyl)-L-glutamine, also called volicitin⁵. And the saliva of spider mites contains an enzyme, β -glucosidase, that induces volatile emissions from lima-bean

plants⁶. But it is not yet known how these 'elicitor' molecules signal to the plants, or how they interact with or complement the plants' own elicitors. Moreover, when the corn and lima-bean plants are attacked these volatiles are even released from undamaged leaves, so presumably a compound from the wound site is transported in the vascular systems of the plant, causing the emission of volatile chemicals from all parts of the plant⁴. Such systemic signalling in response to herbivores is well established⁷. But, except for the polypeptide defence signal called systemin⁸, little is known of the chemical nature of the systemic signals produced in plants by insect attacks.

Many herbivorous insect larvae that have large effects on agriculture are hatched from eggs deposited on the plants at night by nocturnal moths. The larvae feed mainly during daylight hours, which is when the plants release volatiles that attract predatory and parasitic insects. Little is known about volatiles that might be emitted at night.

De Moraes *et al.*¹ now show that tobacco plants under attack from caterpillars produce volatiles at night, as well as by day, and that the nocturnal chemicals are different to the daytime ones. The night emissions,

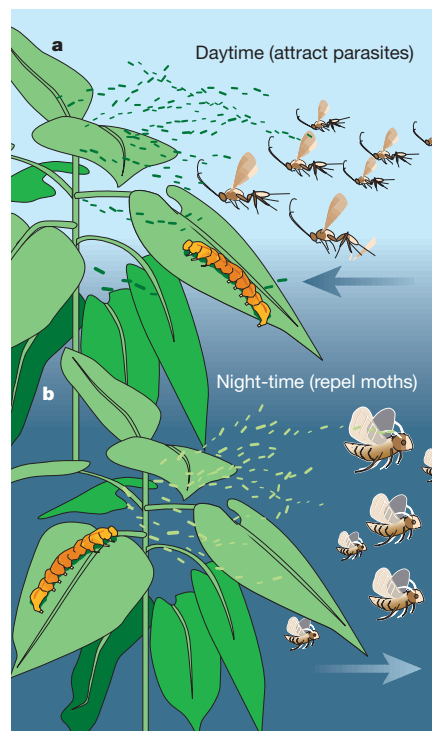


Figure 1 Tobacco plants use different blends of volatile compounds by day and by night, as shown by De Moraes and colleagues¹. a, In daylight, plants under attack by herbivores emit blends that attract parasitic or predatory insects, which destroy the herbivores. b, By night, attacked tobacco plants release volatiles that repel nocturnal pregnant moths, which are looking for somewhere to deposit their eggs. Both responses are beneficial to the plants, and also to the insect species involved.

which are mainly blends of small, unsaturated derivatives of fatty acids, discourage foraging moths (such as *Heliothis virescens*) from laying eggs — which, if they hatched, would soon make a bad situation worse for the plant. So this system benefits both the plants and (presumably unwittingly) the nocturnal moths.

These observations¹ raise questions about the full extent of the volatile language used by plants to communicate with each other and with insects, and the possible relevance of these interactions to ecology and agriculture. If the volatile compounds involved, and the genes needed to produce them, could be tracked down, it might then be possible to develop plants that — even when not under attack — emit blends of defensive volatiles at appropriate times of day and night. Even if the volatiles were

produced at very low levels, the plants might discourage moths from laying eggs at night while attracting predatory insects during the day, significantly minimizing the need for environmentally detrimental pesticides. ■

Clarence A. Ryan is at the Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340, USA.
e-mail: cabudryan@hotmail.com

1. De Moraes, C. M., Mescher, M. C. & Tumlinson, J. H. *Nature* **410**, 577–580 (2001).
2. Dicke, M. et al. *J. Chem. Ecol.* **16**, 3091–3118 (1990).
3. Turlings, T. C. J. et al. *Science* **250**, 1251 (1990).
4. Takabayashi, J. & Dicke, M. *Trends Biochem. Sci.* **1**, 109–113 (1996).
5. Alborn, H. T. et al. *Science* **276**, 945–949 (1997).
6. Mattiacci, L., Dicke, M. & Posthumus, M. A. *Proc. Natl Acad. Sci. USA* **92**, 2036–2040 (1995).
7. Karban, R. & Baldwin, I. T. *Induced Responses to Herbivory* 26–39 (Univ. Chicago Press, 1997).
8. Ryan, C. A. *Biochim. Biophys. Acta* **1477**, 112–121 (2000).

Nanotechnology

Dragging single electrons

Konstantin Likharev

Controlled transfer of single electrons in nanoscale devices allows electric current to be measured in fundamental units of charge and frequency. A new silicon device promises to make these standards more practical.

Accurate measurements of basic physical units such as time, length, mass and electric current are vital for many areas of science and engineering. But the standards are better defined for some of these units than for others¹. For example, standard measurements of frequency (and hence time) use atomic clocks, which are accurate to at least 10^{-14} . Among the electric units, voltage and resistance can be measured to an accuracy of at least 10^{-10} . But although Ohm's law (voltage = current $\times$ resistance) relates electric current to voltage and resistance, measurements of current were, until a few years ago, relatively crude at an accuracy of only about 10^{-6} . On page 560 of this issue², Fujiwara and Takahashi describe the manipulation of a single positive charge carrier, or 'hole', which could potentially increase the accuracy of current measurements.

Fundamental standards of units can be obtained by taking careful measurements in quantum experiments. For example, in atomic clocks, frequency is measured in terms of quantum transitions within atoms. For voltage, measurements with an accuracy of at least 10^{-15} can be made by taking advantage of the Josephson effect, in which current flows by 'tunnelling' through an insulating barrier between two superconductors. Resistance is measured using the quantum limit of the Hall effect — the generation of voltage by a magnetic field acting on a current — but this is more sensitive

to experimental imperfections and seems to be limited to an accuracy of about 10^{-10} . But although voltage and resistance units are

highly accurate, the Josephson and quantum Hall effects cannot be combined to give an accurate measure of current because the experimental conditions and voltage scales for these effects are incompatible.

One way in which current can be measured is to use single-electron devices³. The initial idea⁴ to use these devices for current standards proved not very practical. Eventually, French scientists working in CEA Saclay produced⁵ a more promising device called the single-electron pump (Fig. 1a). This pump is made up of a chain of nanoscale conducting islands, separated from each other by thin insulating barriers. The barriers allow quantum-mechanical tunnelling of electrons, in which the electrons pass through an energy barrier that they would not be able to overcome under classical conditions. Tunnelling is controlled using gate electrodes coupled with each island.

The first gate lowers the Fermi level — the maximum energy of electrons in a solid — of the first island relative to that of the source electrode, creating a potential well and allowing one electron to tunnel through the insulating barrier into the island. Because the island is so small, electrostatic (Coulomb) repulsion prevents additional electrons from following. The next gate lowers the Fermi level of the second island even further, so the electron can tunnel forward into that island. Repeating this procedure sequentially with all the

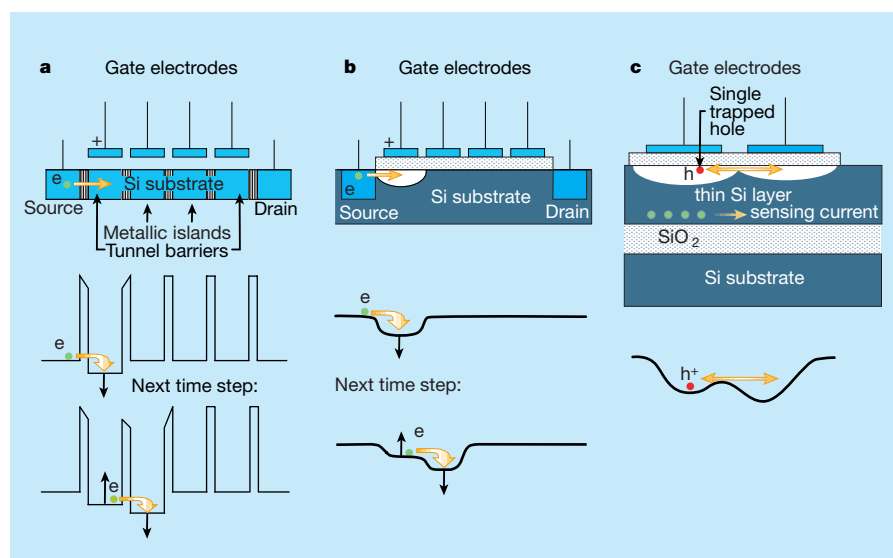


Figure 1 Three devices for transferring individual charge carriers with corresponding potential energy wells shown underneath. Each device has a set of gate electrodes, and applying voltage potentials to the gates allows individual charge carriers to be captured and moved from the source electrode to the drain electrode. a, The 'single-electron pump'⁵ allows an electron to pass from one conducting island to the next by quantum tunnelling through the energy barriers separating the islands. The potential barriers in the pump are not to scale; the real height of about 2 electron volts (eV) is much larger than the potential well's depth (~ 0.001 eV). b, A hypothetical single-electron charged-couple device³ that can drag a single electron in the same way as a water pump but without tunnelling barriers. c, The device made by Fujiwara and Takahashi² that can transfer a hole between two potential wells. The hole's position can be controlled by the 'sensing' electron current, created by the 'field-effect transistor', that flows along the opposite surface of the thin silicon (Si) layer.

islands 'drags' a single electron through the pump. As the current is defined as the passage of charge over unit time, this repetitive process with well-known frequency gives an equally well-known current. Using a six-island circuit, researchers at the National Institute of Standards and Technology⁶ obtained a record current accuracy of 1.5×10^{-8} .

The main practical limitation to using pump standards is not the accuracy, which might still be improved, but rather the low value of the current, which, at about 10^{-12} ampere, is hard to measure. Faster drive frequencies would result in higher currents. But in a pump with insulating tunnel barriers, the drive frequency is limited to a few million cycles per second (MHz), to give the electrons enough time to tunnel through the barriers. Theory indicates⁷ that insulating tunnel barriers are not really needed for ordered single-electron transport in one-dimensional channels. The first experiments of this kind were carried out in 1996 by a group at Cambridge University⁸, which used surface acoustic waves to create a potential profile that moves across the device, dragging the electrons at a rate of up to a few gigahertz (10^9 Hz). But this is a very indirect way to create potential wells, and to my knowledge no current accuracy better than 3×10^{-5} has been reported using this method.

A more promising way³ would be to use nanoscaled charge-coupled devices (CCD). Usually, CCDs⁹ are simple and robust silicon chips such as those at the heart of every camcorder. In a commercial CCD, the size of the potential well dragging the electrons along the chip surface is relatively large — a few micrometres. As a result, the number of electrons in the well is huge (up to a million), and fluctuates strongly even at steady conditions. Scaling the CCD to give a nanometre-sized well might allow Coulomb repulsion to limit the number of electrons present in a way that can be controlled precisely, while still operating at frequencies of up to 1 GHz and beyond (Fig. 1b).

Fujiwara and Takahashi² have taken the first experimental step towards making a single-electron CCD pump. They have created a prototype silicon (Si) device with only two potential wells (Fig. 1c). A 'hole' can be trapped in either of these wells and transferred from one to the other. Each well is coupled to a miniature transistor, which may control a current of electrons within the silicon. Fujiwara and Takahashi use this current to 'sense' the trapped hole and monitor its transfer between the wells.

A practical current standard would not need the electrometers, but would require a source, drain and special external electronics to compensate for random electric fields near the wells³. As such electronics have already been developed for the standard sin-

gle-electron pumps⁶, one may hope to see the first practical CCD current standards in the near future. ■

Konstantin Likharev is in the Department of Physics and Astronomy of the State University of New York, Stony Brook, New York 11794-3800, USA.

e-mail: klikharev@notes.cc.sunysb.edu

1. Cutler, P. H. & Lukas, A. A. (eds) *Quantum Metrology and Fundamental Physical Constants* (Plenum, New York, 1983).

2. Fujiwara, A. & Takahashi, Y. *Nature* **410**, 560–562 (2001).
3. Likharev, K. *Proc. IEEE* **87**, 606–632 (1999).
4. Likharev, K. & Zorin, A. *J. Low Temp. Phys.* **59**, 347–382 (1985).
5. Pothier, H. *et al.* *Physica B* **169**, 573–574 (1991).
6. Keller, M. W. *et al.* *Appl. Phys. Lett.* **69**, 1804–1806 (1996).
7. Averin, D. & Likharev, K. *Nanostructures and Mesoscopic Systems* (eds Kirk, W. & Reed, M.) 283–295 (Academic, Boston, 1992).
8. Shilton, J. M. *et al.* *J. Phys. Cond. Mat.* **8**, L531–L539 (1996).
9. Barbe, D. F. (ed.) *Charge-coupled devices* (Springer, Berlin, 1980).

Cognitive neuroscience

Colour my i's blue

Lynn C. Robertson

Studies of people who perceive colours when they see particular letters or digits are providing help with an old problem — that of whether awareness is needed to 'bind' visual features of an object together.

Most of the information stored in the human mind is below the level of awareness. But what form does this information take? This is an ongoing debate, particularly in the field of visual perception^{1,2}, where it centres around the question of whether the things we see are stored as fragments and features, or as integrated, whole replicas. More precisely, are features of an object (such as shape, colour, brightness and orientation) 'unbound' when we are not aware of them, becoming 'bound' together as a unit only when attention is engaged? The paper on page 580 of this issue³ by Mattingley and colleagues is consistent with a growing body of evidence^{4,5} that attention may be necessary to perceive a unified world. But Mattingley *et al.* address the issue in an intriguing and creative way — by studying a group of people who perceive colour when they see particular 'graphemes' (such as letters or digits).

In these people, presentation of the letter 'a' may evoke the perception of red, for instance, while the letter 'i' may evoke the colour blue. According to individuals who have such 'colour-graphemic synaesthesia', the colour they see is the same whenever the inducing grapheme is seen. These unusual experiences appear to be present from birth and to run in families⁶. There are various forms of synaesthesia, which formally means 'the joining of the senses'. But all 15 of the participants in Mattingley *et al.*'s study³ reported seeing distinct colours when certain graphemes were presented, and they consistently paired a rectangular patch of one particular colour with specific letters or digits.

The authors objectively confirmed these reports by presenting a letter or digit in a colour that was either the same or different as that evoked by the grapheme. They then measured the time it took each person to name the real colour of the stimulus. In

comparison with controls, synaesthetic participants were quicker at naming the colour of the letter when it was the same as that evoked by the stimulus, and slower when the colour of the stimulus was different. This finding has also been reported in single, isolated cases⁷.

Mattingley *et al.* then asked the participants to perform the same task — colour naming — but used a slightly different procedure. Grey graphemes that consistently evoked a particular colour were displayed for a very short amount of time; the participants were then asked to name the colour of a subsequent rectangular patch. When the grey grapheme was presented for 500 milliseconds before the coloured patch, the subjects had enough time to become aware of that grapheme's shape, and the effects observed in the first experiment were replicated. In other words, although the grapheme disappeared before the rectangular patch appeared, the colour evoked by the grapheme in the synaesthetic participants affected how quickly they could name the colour of the patch.

By contrast, when the grey grapheme was preceded by 28 or 56 milliseconds by a 'visual mask', the effects of that grapheme on colour naming disappeared. A visual mask is simply a pattern that reduces the visibility of the stimulus (the grapheme) and does not allow the brain to process that stimulus to the level of conscious awareness. (When the participants were not aware of having seen the grapheme, none of them, whether controls or synaesthetes, could correctly say which grapheme had been presented.) So, when the synaesthetic participants were not consciously aware of the colour-evoking grapheme — when the processing of that stimulus in the brain had not gone beyond a certain point — then the colour they normally saw was not evoked.

But did the visual system actually form



100 YEARS AGO

The remarkable subsidences which have often occurred in and around the town of Northwich, in Cheshire, form the subject of a paper by Mr. T. Ward, recently issued by the Institution of Mining Engineers. The subsidences are chiefly due to mining in the Upper Bed of rock-salt, and the too rapid removal of brine by means of modern pumps. In a natural condition the water in or on the salt-beds becomes saturated with salt and then ceases to dissolve it, but now the brine is continually pumped up in immense quantities, and the fresh water which flows to take its place dissolves the salt pillars which have supported the roof and overlying strata, with the result that there is a depression towards each pumping centre. In almost every case the mines in the Upper Bed of rock-salt are destroyed by water rapidly eroding the salt pillars in this way. Another cause of subsidence is the pumping of brine from off the rock-head, that is, the surface of the Upper Bed of rock-salt... By degrees the town is becoming one of framework buildings, and will, for England, be unique in this respect. The accompanying illustration, which we are enabled to give from Mr. Ward's paper, shows a subsiding house in a street at Northwich.



From *Nature* 28 March 1901.

50 YEARS AGO

John Constable's Clouds. This is more of a philosophical than a meteorological treatise. The author is mainly concerned to demonstrate the influence upon John Constable and, incidentally, also upon Goethe, of Luke Howard's system of cloud classification. There have been other great cloud painters such as Turner and Ruisdael; but Constable's work rests on a knowledge of Howard's system and it is this which makes the dynamical quality of his cloud studies, so suggestive of rapid change, highly interesting to meteorologists. This is surely a striking example of the aid which science can render art...

From *Nature* 31 March 1951.

an early representation of the masked grey grapheme? After all, it would not be surprising that something that was not seen would not influence later processing (colour naming in this case). To tackle this question, Mattingley *et al.* used the same graphemes, visual mask and temporal parameters in another study. Now, though, the coloured patch was replaced by either the same or a different grapheme but in a different case. For example, a lower-case masked grey 'a' was followed by an upper-case unmasked grey 'A'. The task was to name the upper-case letter, which the participants saw clearly.

Although neither synaesthetic participants nor controls consciously saw the masked grey grapheme, all were affected by it. That is, they were quicker at naming the letter 'A' when it was preceded by an 'a' rather than a 'b'. The implication is that although the participants were not consciously aware of the identity of the masked graphemes, their brains were. Nonetheless, the masked graphemes did not evoke synaesthesia. The synaesthetic experience occurred only when the evoking stimulus was consciously perceived, not when it was represented below the level of awareness.

What does all this have to do with the problem of 'binding' visual features of an object into a whole? Neurobiological evidence has shown that separate features of visual information are projected to different cortical regions of the human brain. Colour and shape are separated relatively early during the processing of visual stimuli⁸, and the brain can encode these features without awareness. The new work³ supports this idea of modularity in the human cortex.

It is possible that a flaw in this modular organization accounts for colour-graphemic synaesthesia⁹. Mattingley *et al.*'s results agree with this possibility; it seems that the cortical regions for processing shape and colour are abnormally linked, but only during awareness. The authors' findings are consistent with evidence that 'attention' signals — associated with awareness and generated outside these cortical regions — are required to produce normal binding⁴. For colour-graphemic synaesthetes, 'normal' means seeing a selected grapheme in a particular colour; some degree of awareness is needed to bind colour and grapheme together into one visual object. 'Normal' for other people means seeing the grapheme as the colour it actually is and, again, awareness is likely to be fundamental in binding together colour and shape.

Lynn C. Robertson is in the Office of Veterans Affairs, 150 Muir Road, Martinez, California 94553, and the Department of Psychology, University of California, Berkeley, California 94720, USA.

e-mail: lynnrob@socrates.berkeley.edu

1. Treisman, A. *Phil. Trans. R. Soc. Lond. B* **353**, 1295–1306 (1998).
2. Wojciulik, E. & Kanwisher, N. *Vis. Cogn.* **5**, 157–181 (1998).
3. Mattingley, J. B., Rich, A. N., Yelland, G. & Bradshaw, J. L. *Nature* **410**, 580–582 (2001).
4. Friedman-Hill, S., Robertson, L. C. & Treisman, A. *Science* **269**, 853–855 (1995).
5. Robertson, L. C., Treisman, A., Friedman-Hill, S. & Grabowecy, M. *J. Cogn. Neurosci.* **9**, 295–317 (1997).
6. Baron-Cohen, S. & Harrison, J. E. (eds) *Synaesthesia: Classic and Contemporary Readings* (Blackwell, Cambridge, MA, 1997).
7. Dixon, M. J., Smilek, D., Cudahy, C. & Merikle, P. M. *Nature* **406**, 365 (2000).
8. Desimone, R., Stanley, J. S., Moran, J. & Ungerleider, L. G. *Vision Res.* **25**, 441–452 (1985).
9. Baron-Cohen, S., Harrison, J., Goldstein, L. H. & Wyke, M. *Perception* **22**, 419–426 (1993).

Palaeontology

Chinese salamanders tell tales

Robert Carroll

A palaeontological treasure trove of over 500 specimens will help in understanding the evolutionary history of salamanders and their relationships with other amphibians.

On page 574 of this issue¹, Gao and Shubin report the discovery of hundreds of superbly preserved salamander skeletons from a 150-million-year-old pond deposit in China. The specimens include both larval and adult stages. They will help to answer questions about the evolutionary relationships and geographical distribution of the 10 living families of salamanders, as well as about the ancestry of salamanders — the order Caudata (urodeles) — as a whole. The urodeles are one of three orders that comprise the modern amphibians, the others being frogs and a

limbless group known as the caecilians. The age of this deposit is based on radiometric dating and association with other fossils common to the Upper Jurassic (see Fig. 1 for timescale).

More than 500 specimens were collected from an area of less than 10 m² that had been buried by a volcanic eruption. The specimens include two species whose growth stages illustrate the main life-history strategies of modern salamanders. One exhibits the pattern of metamorphosis common to many living amphibians: aquatic larvae possessing external gills, with the gills being

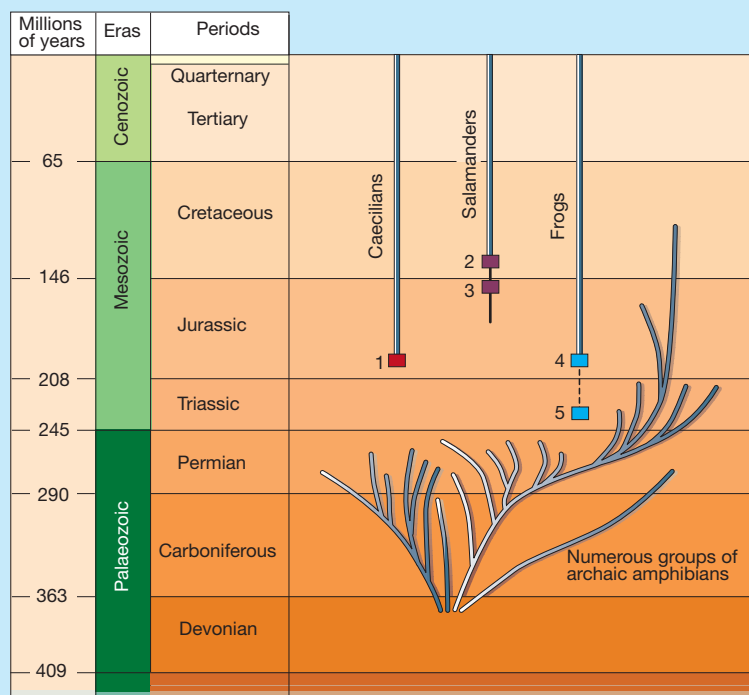


Figure 1 Geological timescale showing the overall pattern of amphibian evolution, and the time of occurrence of the fossils discussed by Gao and Shubin¹ and shown in Fig. 2. 1, *Eocaecilia micropodia*. 2, *Valdetrion gracilis*. 3, The newly described¹ Chinese species. 4, *Prosalirus bitis*. 5, A Lower Triassic amphibian, close to the ancestry of frogs.

lost in larger, more fully ossified specimens that are presumed to be terrestrial. The second species, like members of many modern salamander families, was apparently neotenic — that is, it reached sexual maturity without metamorphosing into a fully terrestrial form. That some examples of this species had reached maturity can be seen from the high level of ossification of the bones in the wrists and ankles. But the gill

supports are also ossified, indicating that the creature still relied on external gills for aquatic respiration.

The fossils are immediately recognizable as salamanders from their body and limb proportions, as well as from details of the skull anatomy. They also share with modern salamanders a unique aspect of limb development, in which the two most anterior elements of the distal row of bones in the wrist

and ankle are fused to form a single bone, known as the basal commune, that supports the first two digits. No group of vertebrates other than salamanders has this specific configuration.

The closest relationships of the Chinese specimens with living salamanders lie within the most primitive families, the Hynobiidae and Cryptobranchidae, which together constitute the superfamily Cryptobranchoidea. The Chinese fossils and cryptobranchoids share the retention of several bones that are lost in more advanced salamanders (including the lacrimal bone, which bears the lacrimal duct, and a separate angular bone in the lower jaw). But the fossil specimens appear to be even more primitive in their retention of a separate coronoid bone in the adult and the presence of more than three caudal vertebrae bearing ribs. Given the close affinities of the Jurassic species with the cryptobranchoids, Gao and Shubin argue that the initial radiation of all later salamander lineages occurred in Asia, before they spread to Europe and North America.

The primitive position of the Chinese species is evident from Gao and Shubin's phylogenetic analysis, which combines skeletal data from fossils and modern families with molecular evidence from living amphibians. But the consensus trees provide little resolution of the specific pattern of interrelationships of the more advanced families. Further study of these and other Mesozoic salamanders from China should allow the sequences in which the families diverged to be more reliably established.

The ancestry of salamanders remains highly contentious, and the newly discovered fossils should help here, too. It is generally accepted that salamanders shared

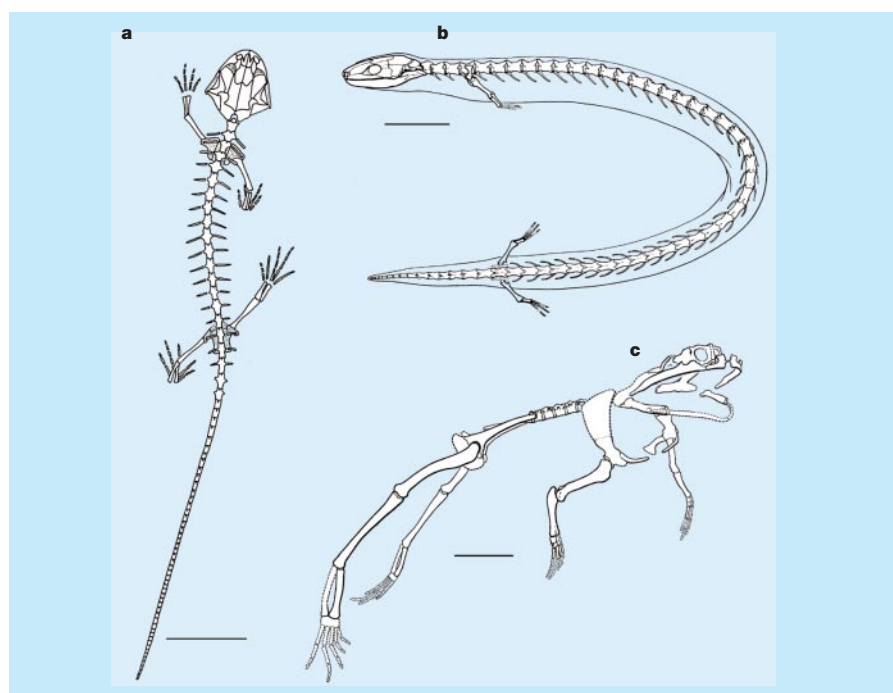


Figure 2 Comparison of the body proportions and limb structures of a fossil salamander, caecilian and frog. a, *Valdetrion gracilis*⁸, a salamander from the Lower Cretaceous of Spain. This species broadly resembles the Upper Jurassic salamander, described by Gao and Shubin¹, which underwent metamorphosis from an aquatic larval stage to a terrestrial adult. b, A caecilian (*Eocaecilia micropodia*) and, c, a frog (*Prosalirus bitis*) from the Lower Jurassic of Arizona. The distinct anatomy and way of life of Jurassic salamanders, caecilians and frogs, and the presence of a putative frog ancestor from the Lower Triassic, suggest that separate ancestral lineages for all three orders should be sought among Palaeozoic amphibians. Scale bars are 1 cm. Parts a, b and c are reproduced from refs 8, 5 and 4, respectively.

with frogs and caecilians (which have greatly elongated bodies but lack limbs, and are restricted to the wet tropics) an immediate common ancestry that was distinct from that of all other terrestrial vertebrates. But the various hypotheses about the Palaeozoic ancestry of salamanders differ considerably^{2,3}.

The Chinese fossils and living cryptobranchoids show a host of features that indicate a long period of evolution after they had diverged from the ancestors of frogs and caecilians. The limb proportions of primitive salamanders are similar to those of most Palaeozoic amphibians, but they show no evidence of the jumping habit of frogs or the snake-like body of caecilians that were already established by the Lower Jurassic^{4,5} (Fig. 2). Whereas frogs had evolved a highly derived tadpole by the Lower Cretaceous, with a distinctive filtering apparatus for feeding on microscopic plant material, salamanders retained gradual metamorphosis without a distinct change in diet between the larvae and adults.

The high degree of adaptive and structural divergence of the modern amphibian orders does not necessarily mean that they did not ultimately have a common ancestry. But it does point to the importance of establishing the specific ancestry of each of the three lineages, and how they may be related to the great diversity of Palaeozoic amphibians.

In most aspects of their body proportions and feeding apparatus, primitive living salamanders resemble the Palaeozoic branchiosaurs⁶ — small, immature amphibians with external gills. Like modern salamanders, many branchiosaurs were neotenic; this is a lifestyle that is not possible for frogs because they cannot reproduce before metamorphosis⁷.

We have no fossil from the Upper Permian or Triassic to link the essentially modern salamanders of the Jurassic with any putative Palaeozoic ancestors. But the Chinese specimens now provide a way to compare the two — including not only the adult anatomy, but also the generally untapped information from larval stages and patterns of development that can also be studied in branchiosaurs.

Robert Carroll is at the Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal H3A 2K6, Canada.

e-mail: robertc@shared1.lan.mcgill.ca

1. Gao, K.-Q. & Shubin, N. H. *Nature* **410**, 574–577 (2001).
2. Milner, A. *Herpetol. Monogr.* **7**, 8–27 (1993).
3. Laurin, M. & Reisz, R. in *Amniote Origins* (eds Sumida, S. S. & Martin, K. L. M.) 9–59 (Academic, San Diego, 1997).
4. Shubin, N. H. & Jenkins, F. A. *Nature* **377**, 49–52 (1995).
5. Jenkins, F. A. & Walsh, D. M. *Nature* **365**, 246–249 (1993).
6. Boy, J. A. & Sues, H.-D. in *Amphibian Biology* Vol. 4, *Palaeontology: The Evolutionary History of Amphibians* (eds Heatwole, H. & Carroll, R. L.) 1150–1197 (Surrey Beatty, Chipping Norton, New South Wales, 2000).
7. Wassersug, R. J. *Am. Zool.* **15**, 405–417 (1975).
8. Evans, S. E. & Milner, A. R. *Phil. Trans. R. Soc. Lond. B* **351**, 627–646 (1996).

Cancer

Chromosome defects in the colon

David M. Livingston

A tumour-suppressor protein known as adenomatous polyposis coli malfunctions in many colon cancers. So too does chromosome segregation during cell division. These features may somehow be connected.

The cells of many colon cancers have the wrong number of chromosomes, as well as defects in a protein called adenomatous polyposis coli (APC). Writing in *Nature Cell Biology*, Fodde and colleagues¹ and Kaplan and co-workers² provide surprising evidence for a cause-and-effect relationship between these two phenomena. The key to the puzzle is that, as well as having numerous other roles, the functional APC protein docks at chromosomal regions called kinetochores during cell division. Failure to do this seems to result in defects in the segregation of chromosomes into two daughter cells during cell division, as well as abnormal chromosome numbers.

The APC protein is a tumour suppressor; in other words, it normally contributes to the suppression of cancerous cellular characteristics. According to doctrine, APC's tumour-suppressing function is intimately linked to its participation in the efficient turnover of

β -catenin. This protein is a key component of the Wnt signalling pathway — a multistep process that is involved in controlling development and cell proliferation. The best evidence for this view of APC's function is that colon cancer cells either have mutations in APC that are associated with accumulation of β -catenin, or have mutations in β -catenin that promote its own accumulation. In the latter case, the colon tumour cells have normal APC.

But APC has other functions, too. It is also involved in regulating the functions of certain networks of molecules that make up the cellular cytoskeleton. For example, it interacts with the ends of growing microtubules (a type of cytoskeletal filament), a function that is mediated by its carboxy-terminal region^{3–5}. This portion of APC is also, at least in part, dedicated to interactions with EB1 — a protein that also interacts with microtubules, including those associ-

ated with the spindle (the chromosome-segregating apparatus) and centrosomes (organelles that produce the spindle)^{6–10}.

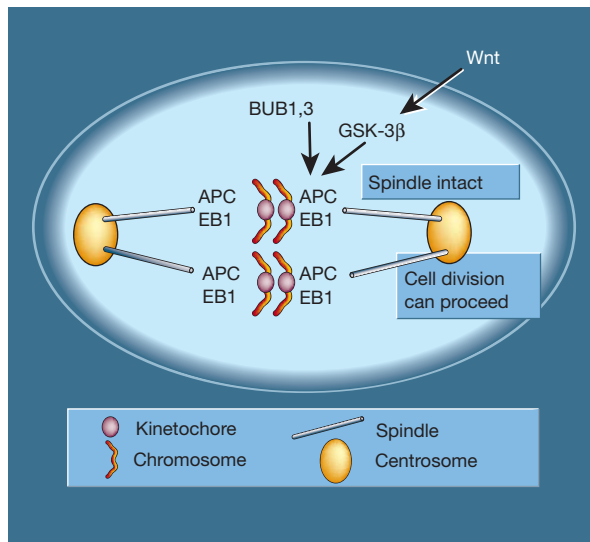
Given the link between APC and microtubules, and the involvement of microtubules in the spindle, Fodde *et al.*¹ and Kaplan *et al.*² wondered whether there is a relationship between the defects in chromosome segregation seen in colon cancer cells bearing mutations in APC¹¹, and the APC–microtubule connection. They show that, in cells that are dividing, APC is located at the ends of spindle microtubules that associate with kinetochores (Fig. 1) — the chromosomal regions to which the spindle attaches. Indeed, APC itself interacts with kinetochores, and this depends on the presence of an intact spindle apparatus. The authors also generated mutations in APC in mouse embryonic stem cells, which are embryonic cells with the potential to develop into many other cell types. The mutant stem cells, too, had aberrant chromosome numbers and underwent defective chromosome segregation — abnormalities that were not detected in controls. So there seems to be a direct link between normal APC function and chromosome segregation.

A yeast analogue of the mammalian APC-binding protein EB1 is known to be involved in the function of the spindle and in chromosome segregation^{12–16}. So Fodde *et al.* searched for, and found, interactions between EB1 and kinetochores in mouse embryonic stem cells. They went on to show that disruption of the formation of an APC–EB1 complex was associated with abnormal spindle formation and incorrect chromosome numbers. So EB1 probably provides at least part of the molecular connection between APC and normal spindle function.

But how might APC influence processes at the kinetochore–spindle interface? The assembly of a spindle is crucial for correct cell division, and a 'spindle-assembly checkpoint' ensures that cells do not attempt to segregate their chromosomes until a proper spindle is in place. Neither group of authors detected a role for APC in the spindle checkpoint. But Kaplan *et al.* do provide evidence that two enzymes that are involved in the checkpoint, BUB1 and BUB3, probably associate with APC. Moreover, *in vitro* these enzymes efficiently added phosphate groups to APC.

Even more telling, this reaction was enhanced markedly by previous phosphorylation of APC by glycogen synthase kinase-3 β , a key participant in the Wnt signalling pathway. So it appears that APC communicates with 'professional' spindle-checkpoint proteins, and might be influenced by them in a Wnt-dependent manner. Given these observations, one feels compelled to ask whether APC has finally grown beyond its distinguished origins in the Wnt pathway

Figure 1 Colon cancer cells have defects in the adenomatous polyposis coli (APC) protein, as well as aberrant segregation of chromosomes during cell division. Fodde *et al.*¹ and Kaplan *et al.*² show that these phenomena are linked. They find that APC and an APC-binding protein, EB1, interact with kinetochores (the chromosomal regions to which the chromosome-segregating machinery, the spindle, attaches) and with the ends of the spindle attached to the kinetochores. Failure to do this results in chromosome-segregation defects. It is uncertain how this might come about, as APC does not contribute in a major way to the spindle-assembly checkpoint, which monitors whether an intact spindle is formed before allowing chromosome segregation to proceed. But APC is phosphorylated by BUB1 and BUB3, which are components of the spindle checkpoint. APC was previously known to participate in the Wnt signalling pathway, which is involved in development and cell proliferation. The activation of this pathway leads to the phosphorylation of APC by glycogen synthase kinase-3 β (GSK-3 β). It now seems that there is crosstalk between these pathways: BUB1 and BUB3 phosphorylate APC more effectively when APC is already phosphorylated by GSK-3 β .



and assumed a second career in the control of cell division. Similarly, it appears that the influence of Wnt transcends its original role of regulating β -catenin.

But if cells that lack normal APC function have an intact spindle checkpoint, how do they survive their characteristic defect in chromosome segregation? The simplest answer would be that APC deficiency is not the sole contributor to chromosome instability in cells in which an APC defect is clinically significant (colon cancer cells). Further mutations in genes such as *ras*, *c-myc* and *p53* may also be required to silence the compensatory or checkpoint systems that are normally dedicated to eliminating cells with defective chromosome segregation.

However, Fodde *et al.* and Kaplan *et al.* found embryonic stem cells, with engineered mutations only in APC, that had abnormal numbers of chromosomes. Perhaps there are as-yet-undefined pathways for detecting aberrant chromosome segregation that do not work efficiently in certain stem cells. Indeed, embryonic stem cells are known to have defects in the 'G1' checkpoint¹⁷. This checkpoint is not involved in controlling chromosome segregation, but the fact that embryonic stem cells have defects in this pathway makes one wonder whether they also have intrinsic problems in monitoring chromosome segregation.

Colon cancer cells share at least one characteristic with the stem cells from which they are derived: persistent self-renewal. Is the development of this feature in differentiated epithelial cells, such as colon cells, sometimes associated with inefficient checkpoint

control? If so, loss of function of APC, or of other proteins dedicated to normal control of cell division, in a colon cell with stem-cell-like properties might evoke features of genome instability without the need for further genetic perturbations.

Whatever the case, the story of APC's tumour-suppressing function has grown richer and more complex. In particular, a role for this protein in controlling the integrity of the genome must be added to the latest edition of the text.

David M. Livingston is in the Departments of Genetics and Medicine, Harvard Medical School, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA.
e-mail: david_livingston@dfci.harvard.edu

1. Fodde, R. *et al.* *Nature Cell Biol.* **3**, 433–438 (2001).
2. Kaplan, K. B. *et al.* *Nature Cell Biol.* **3**, 429–433 (2001).
3. Munemitsu, S. *et al.* *Cancer Res.* **54**, 3676–3681 (1994).
4. Smith, K. J. *et al.* *Cancer Res.* **54**, 3672–3675 (1994).
5. Deka, J., Kuhlmann, J. & Muller, O. *Eur. J. Biochem.* **253**, 591–597 (1998).
6. Su, L. K. *et al.* *Cancer Res.* **55**, 2972–2977 (1995).
7. Tiauer, J. S. & Bierer, B. E. *J. Cell Biol.* **149**, 761–766 (2000).
8. Berrueta, L. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 10596–10601 (1998).
9. Berrueta, L., Tirnauer, J. S., Schuyler, S. C., Pellman, D. & Bierer, B. E. *Curr. Biol.* **9**, 425–428 (1999).
10. Mimori-Kiyosue, Y., Shiina, N. & Tsukita, S. *Curr. Biol.* **10**, 865–868 (2000).
11. Lengauer, C., Kinzler, K. & Vogelstein, B. *Nature* **386**, 623–627 (1997).
12. Beinbauer, J. D., Hagan, I. M., Hegemann, J. H. & Fleig, U. *J. Cell Biol.* **139**, 717–728 (1997).
13. Muhua, L., Adams, N. R., Murphy, M. D., Shields, C. R. & Cooper, J. A. *Nature* **393**, 487–491 (1998).
14. Schwartz, K., Richards, K. & Botstein, D. *Mol. Cell Biol.* **8**, 2677–2691 (1997).
15. Korinek, W. S., Copeland, M. J., Chaudhuri, A. & Chant, J. *Science* **287**, 2257–2259 (2000).
16. Lee, L. *et al.* *Science* **287**, 2260–2262 (2000).
17. Aladjem, M. I. *et al.* *Curr. Biol.* **8**, 145–155 (1998).

Daedalus

Natural currents

The electrochemical industry is largely based on the electrolysis of brine, giving sodium (or sodium hydroxide) and chlorine (or hydrogen chloride). The vast advances in this process have not been matched by the creation of the electricity needed, which is still largely based on steam equipment. This will soon collide with the green ambitions of the British government.

Daedalus has therefore been looking at natural sources of high currents at low voltage (the electrochemical ideal). He notes the Gulf Stream, for example, which is largely responsible for warming Britain. As it sweeps between Iceland and the north of Scotland, it cuts across the vertical component of the Earth's magnetic field. This flux-cutting must generate about 10 volts, even if the Gulf Stream is quite sluggish. But Daedalus recalls that many surface ocean currents are superficial affairs, generated by the prevailing winds. The Gulf Stream is more serious, part of the North Atlantic Conveyor; even so, there must be a return flow, quite possibly in some other direction.

So Daedalus plans a big polypropylene sheet, stretching at depth from the north of Scotland (Dounreay is usefully industrialized already) to off Reykjavik in Iceland. The polyalkanes are less dense than sea water, and it should be possible to give the sheet a density such that it would float stably at the desired depth, or else it could be floated stably on the denser return flow. Even with just a few volts, the huge expanse of ocean above and below the sheet will have such a low internal resistance that a vast current would be available between the terminations of the sheet at Dounreay.

The power for the project would ultimately come from the Sun, or the slowing of the Earth–Moon system. (Daedalus, a lazy fellow, rather likes the idea of a 25-hour day). And, as for any natural power source, efficiency is quite unimportant. The more sheet, the more available current. If cold weather began to indicate the successful slowing or stoppage of the Gulf Stream, there would be plenty of time to reduce the drawn power, or otherwise modify the set-up.

Meanwhile sea water, almost entirely brine anyway, would be electrolysed almost automatically at Dounreay, and the chemical industry could breathe again. Other electrolytic industries, such as the manufacture of aluminium, now in places chosen to exploit hydroelectric power, could also be drawn to the site. **David Jones**

Obituary

Leonard Mandel (1927–2001)

Leonard Mandel, one of the founding fathers of quantum optics, died at his home on 9 February at the age of 73. Our generation of optical scientists grew up learning from Mandel's work: quite literally, he showed us the light by opening up many new avenues of research. He was an elegant experimentalist and a celebrated theorist. Echoes of his unmistakable slow but clear voice will continue to remind us of the contributions of this remarkable man.

Mandel was born in Berlin, but moved to Britain when he was a boy. He was educated at the University of London, studying cosmic rays for his PhD, which was awarded in 1951. After a short spell in industry, Mandel took up an academic post at Imperial College, London. It was the investigation of cosmic rays that eventually led him to discover what is today known as Mandel's photon-counting formula, which relates the statistical properties of arbitrary optical fields to those of photoelectrons detected experimentally. Mandel worked throughout his career to perfect the photon-counting technique for studying a variety of optical fields, from the thermal light of a candle to sophisticated quantum sources. In the 1960s, this work enabled physicists to understand the statistical properties of lasers operating under different pumping conditions. Such studies gave support to the quantum theory of the laser.

Mandel performed his first major experiment at Imperial, its simplicity in design becoming one of his trademarks. The result, typically, was counterintuitive. The experiment showed how interference between independent photon beams can be observed and that such a possibility is permitted by quantum physics. His demonstration of two-photon interference followed. This was an experiment that laid the foundation for later work on quantum entanglement, a concept of state superposition first introduced by Erwin Schrödinger.

In 1964 Mandel settled permanently in Rochester, New York, having been persuaded to move there by Emil Wolf. Wolf and Mandel wrote several notable papers together, and also organized the Rochester Conferences on Quantum Optics, which laid the foundations of the field and continue to this day. It was at one of the early conferences that debate really began to heat up on the need for a quantum theory of light, and Mandel started to devote much of his research



Pioneer in quantum approaches to light

to searching for the subtle differences between the classical and quantum descriptions of light.

Applying a mixture of theory and experiment, Mandel and colleagues revealed several uniquely quantum-mechanical properties of light. In 1977, for instance, his group observed the 'photon anti-bunching' effect, in which photons from a single two-level atom illuminated by a resonant laser can never be emitted in pairs or more, providing confirmation of a 'quantum jump'. This was followed by the demonstration of sub-poissonian photon statistics, which mean that fluctuations in the photon number of an optical field are smaller than would be expected for random events, such as the fall of raindrops. These were the first reported observations of 'non-classicality' requiring a full quantum description of light, and they changed our perspective on light.

Photon interference was the hallmark of Mandel's work. In the 1980s his group carried out a series of ground-breaking yet ingeniously simple experiments to demonstrate the quantum entanglement of photons and, subsequently, the non-locality of quantum mechanics through violations of Bell's inequalities in the Einstein–Podolsky–Rosen paradox. These studies set the ball rolling for a large number of experiments world wide dealing with the foundations of quantum mechanics, and for newer developments such as quantum teleportation, quantum encryption, and cryptography for quantum-information processing.

As he entered his seventh decade, Mandel was still as productive as ever in addressing some of the fundamentals of

physics. One of the problems that bothered Paul Dirac, the founder of quantum electrodynamics, was how to define the phase of a quantized electromagnetic field: phase is crucial in understanding all interference phenomena. The answer remained elusive until the early 1990s, when Mandel introduced the concept of measurable phase, and actually measured the phase characteristics of a quantized field.

In this same period he demonstrated the non-existence of a pilot wave, which is at the heart of de Broglie's wave interpretation of quantum mechanics. And in another conceptually straightforward but mind-boggling experiment, Mandel and his students demonstrated the complementarity of light — that is, its behaviour as either a wave or a particle in different circumstances. It was, of course, known that placing a detector in its path could change the behaviour of light. But Mandel found that an experimental set up that merely has the possibility of making a measurement, even if only in principle, could bring that change about: no actual measurement is required.

Mandel was a private but generous person. Discussions with him were invariably inspiring, often leading to fresh ideas, as we ourselves can attest. Although his voice was weakened by illness, Mandel's mind remained as sharp as ever. Only a few months before his death, he and his colleagues published a paper on squeezing the quantum noise in atomic spin, an approach which may lead to entangled states of atoms.

As well as being a world-class researcher, Mandel was a great teacher. Many of his 39 PhD students became leaders in their fields: Jeff Kimble, for instance, with whom Mandel discovered the photon anti-bunching effect, is working at the forefront of the new field of quantum information science, which is based on the concept of quantum entanglement. The field that Mandel opened up, and to which he made so many contributions, has begun to have an impact on the information age in which we now live. His legacy is to have provided the inspiration for the next generation of researchers pursuing a deeper understanding of the fundamental laws of nature and their applications.

G. S. Agarwal and Z. Y. Ou

*G. S. Agarwal is in the Physical Research Laboratory, Ahmedabad 380 009, India.
e-mail: gsa@prl.ernet.in*

*Z. Y. Ou is in the Department of Physics, Indiana University–Purdue University, 402 North Blackford Street, Indianapolis, Indiana 46202, USA.
e-mail: zou@iupui.edu*

Pattern and intensity of physical activity

Keeping moderately active is the best way to boost total daily energy expenditure.

A low level of physical activity is a pervasive feature of our modern lifestyle. Attempts to counteract the negative effects of sedentary living include introducing bouts of high-intensity exercise, but I show here that it can be more effective to increase the amount of time spent on activities of moderate intensity while reducing periods of inactivity during waking hours.

Physical inactivity may be associated not only with being overweight but also with disorders such as coronary heart disease or type-2 diabetes¹, which may be prevented by exercising regularly². Consequently, sedentary people often include spells of vigorous exercise in an otherwise inactive and unhealthy lifestyle. I have tested whether this is the most effective way to enhance people's total daily activity levels by determining how these are influenced by exercise duration and intensity.

Healthy, non-obese adults (14 women and 16 men; 27 ± 5 yr; 24.1 ± 2.3 kg m⁻²) participated in the study³. Total energy expenditure was measured by monitoring their metabolism with doubly labelled water over a two-week period, the optimal observation interval for the biological half-lives of the isotopes. The physical activity level (PAL) is defined as the factor by which total energy expenditure exceeds resting energy expenditure, measured during an overnight stay in a respiration chamber.

To determine the amount of body movement over 1-min intervals, I recorded movements over the first of the two weeks with a 'Tracmor' portable motion sensor. This consists of a body-fixed triaxial accelerometer and a data unit for on-line registration, processing and storage of acceleration signals. The triaxial accelerometer ($50 \times 30 \times 8$ mm; 16 g) consists of three uniaxial piezoresistive accelerometers and registers accelerations in three orthogonal directions as the body moves. Acceleration signals are processed to obtain the sum of the rectified and integrated acceleration curves from all three measurement directions. The integration time is set at 1 min and the final output is expressed as counts per minute (c.p.m.)⁴.

Using data from Tracmor recordings in combination with activity details recorded in diaries, three activity categories were defined: low intensity (lying, sitting and standing; Tracmor output under 1,000 c.p.m.), moderate intensity (walking and cycling; output 1,000–3,000 c.p.m.); and high intensity (housework, gymnastics and sport; output greater than 3,000 c.p.m.).

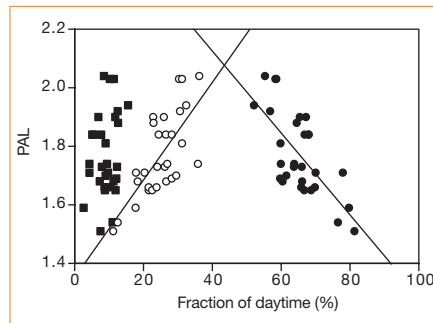


Figure 1 Physical activity levels (PAL) as a function of the fraction of daytime hours spent by 30 healthy subjects (with body-mass indices within the normal range) on activities of low (filled circles), moderate (circles) and high (squares) exercise intensity. Linear regressions are shown for low- and moderate-intensity activities: subjects spending more time on moderate-intensity exercise and less on low-intensity activity can improve their PAL values. Time spent on high-intensity exercise does not appear to influence total energy expenditure.

Calculated PAL values fell in the range 1.51–2.04 for all subjects, apart from one woman whose PAL value was 2.57; her data were excluded as an outlier. Three subjects reported not wearing the Tracmor while performing intensive sports activities, so the time spent on high-intensity activity by these subjects was corrected accordingly.

In a multiple regression analysis with the fraction of time spent on activities of moderate and high intensity, only moderate-intensity activity came out as a significant predictor of PAL ($r^2 = 0.51$, $P < 0.0001$). Subjects spending relatively more time on moderate-intensity activity and therefore less on low-intensity activity had a higher PAL value (Fig. 1). There was, however, no relation between PAL value and the time spent on just high-intensity activity, presumably because this was limited by its nature to being relatively short.

In the general population, PAL ranges between 1.2 and 2.2–2.5. At PAL values of about 2.5, subjects have problems main-

taining energy balance and may lose weight; PAL is about 1.5 or 2.1 for sedentary or very active people, respectively⁵. My results indicating that short periods of vigorous activity do not have much impact on PAL in the normal population are borne out by studies on obese patients: adding exercise to an energy-restricted diet does not further increase weight loss^{6,7} because the costs of the extra activity are compensated by a reduction of energy spent on physical activity outside the training sessions⁸.

My results show, however, that the proportion of time distributed between activities of low and moderate intensity is what influences the total energy expenditure and so determines the value of PAL. Subjects wanting to increase their metabolic rate should exchange low-intensity activities such as sitting in front of a screen for moderate-intensity activities such as walking or cycling. Moderate-intensity activities are better tolerated than high-intensity activities, especially by the middle-aged or obese. An improvement in overall activity levels should bring important health benefits as subjects expend more energy and reduce their risk of running into positive energy balance and gaining weight.

Klaas R. Westerterp

Department of Human Biology, Maastricht University, 6200 MD Maastricht, The Netherlands
e-mail: k.westerterp@hb.unimaas.nl

- Booth, F. W., Gordon, S. E., Carlson, C. J. & Hamilton, M. T. *J. Appl. Physiol.* **88**, 774–787 (2000).
- Grundy, S. M. *et al. Med. Sci. Sports Exerc.* **31** (suppl.), 502–508 (1999).
- Westerterp, K. R. *et al. Br. J. Nutr.* **76**, 785–795 (1996).
- Bouten, C. V. C., Verboeket-van de Venne, W. P. H. G., Westerterp, K. R., Verduin, M. & Janssen, J. D. *J. Appl. Physiol.* **81**, 1019–1026 (1996).
- Westerterp, K. R. *Am. J. Clin. Nutr.* **68** (suppl.), 970–974 (1998).
- Ballor, D. L. & Poehlman, E. T. *Eur. J. Appl. Physiol.* **71**, 535–542 (1995).
- Garrow, J. S. & Summerbell, C. D. *Eur. J. Clin. Nutr.* **49**, 1–10 (1995).
- Kempen, K. P. G., Saris, W. H. M. & Westerterp, K. R. *Am. J. Clin. Nutr.* **62**, 722–729 (1995).

Forensic palaeontology

The *Archaeoraptor* forgery

The *Archaeoraptor* fossil was announced as a 'missing link' and purported to be possibly the best evidence since *Archaeopteryx* that birds did, in fact, evolve from certain types of carnivorous dinosaur¹. It reportedly came from Early Cretaceous beds of China that have produced other spectacular fossils transitional

between birds and extinct non-avian dinosaurs^{2,3}. But *Archaeoraptor* was revealed to be a forgery in which bones of a primitive bird and a non-flying dromaeosaurid dinosaur had been combined^{4–6}. Here we use high-resolution X-ray computed tomography (CT)⁷ to determine the nature and extent of the forgery, as well as how it was built, by imaging the fracture pattern and distribution of materials through the entire specimen.

The *Archaeoraptor* specimen, which was reportedly collected from the Early Creta-

ceous Jiufotang Formation of Liaoning, was smuggled out of China and later sold in the United States on the commercial market⁴. Before repatriation⁵, the specimen, which was alleged to contain a single complete skeleton, was brought to us for CT analysis.

The slab had obviously been reassembled from numerous broken pieces, and the skeleton presented a unique combination of features, including teeth, feathers, a more flight-capable wing than *Archaeopteryx*, and the tail of a non-flying dromaeosaur. The entire bone-bearing portion of the slab was scanned in 422 consecutive 1-mm-thick slices, spaced at 0.9 mm (for the full dataset and description of scanning protocols, see supplementary information).

Computed tomography alone cannot distinguish genuine and fraudulent mistakes, but it provides criteria for evaluating whether pairs of adjacent pieces lie in natural relation to one another. These include comparison of adjacent fracture-face geometries, thickness and density of adjacent pieces, cross-cutting relationships among four generations of fractures, the distribution of grout, and continuity across fractures of bones, natural moulds of bones, and invertebrate burrows. The textural signature of the grout is its relatively low density, inclusion of air bubbles and metallic particles, a unique fracture pattern, and dark green fluorescence under ultraviolet light. Overlapping relationships among blobs of grout and different pieces, which can be observed in CT cross-section, reveal the general sequence of assembly.

The *Archaeopteryx* slab was built in three layers. The top 'layer' is a heterogeneous mosaic of 88 separate pieces, some of which contain bone (Fig. 1). The bottom layer is a single, unbroken plate of shale used for backing. Between the two is a layer of grout, which was spread over the face of the backing plate and between ruptured edges of the top pieces, as they were set into place. Many Liaoning fossils have undergone this same style of repair.

The top layer was built in three phases. The first to be assembled were 23 pieces that lie together in natural relationships and contain the anterior half of an articulated bird skeleton (to be described as a new species; X. Xu *et al.*, manuscript in preparation). In the next phase, 26 pieces containing bones and a few contiguous pieces were sequentially articulated against the rear half of the skeleton to 'complete' it. None of these pieces preserves evidence of a natural attachment to the bird skeleton pieces, and in only a few cases are these pieces naturally associated with one another. Also, the 'paired' tibiae/fibulae and feet are split parts and counterparts from a single side, positioned as if they were right and left.

The third phase of construction involved placement of 39 'shims'. The shims lack

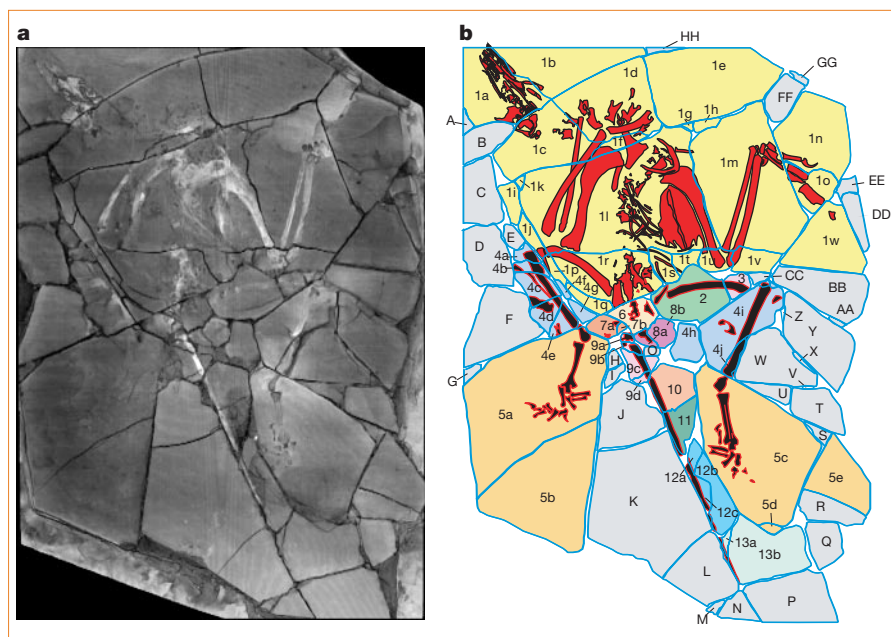


Figure 1 Two computer-generated models of the face of the *Archaeopteryx* slab as it was presented for computed tomography (CT) scanning on 29 July 1999. The specimen was scanned from top to bottom in planes perpendicular to this view. **a**, Volumetric model generated from superimposed CT slices. Light grey, bones; dark grey, slab materials; black, air. **b**, Map of the slab face, colour-coded to indicate the nature of the associations between its 88 constituent pieces. Red, associated bird bones; black, unverifiable 'attached' bones. 1a–w, Associated pieces lying in a natural position; 2, 3, 'left' femur; 4a–j, 'right' and 'left' tibia/fibula (piece and counterpart); 5a–e, 'right' foot/ankle (piece and counterpart); 6, 7a,b, 8a–c, bone fragments; 9a–d, 10, 11, 12a–c, 13a,b, dromaeosaur tail pieces; A–HH, shims. See <http://www.ctlab.geo.utexas.edu/pubs/> for the CT data set and further information.

bone, are usually separated from the other pieces by broad seams of grout, and preserve no evidence of a natural association with any piece in the first two groups. Shims were probably added at various times during assembly to secure the secondary bones into position around the bird skeleton and to make the slab more presentable.

In addition to the discrepancies revealed by the CT information, the tail of a non-volant dromaeosaur also presented character conflicts with the ornithurine features of the bird skeleton and with the foot, which lacks the characteristic enlarged second claw of dromaeosaurids⁸. We found no evidence that the left femur, the split tibiae/fibulae, the split foot, or the tail belonged to the same individual or to the same species.

We conclude that *Archaeopteryx* represents two or more species and that it was assembled from at least two, and possibly five, separate specimens. Additional work in China verified that the tail is from an entirely different specimen, which has been described as a new species of dromaeosaur⁹.

Sadly, parts of at least two significant new specimens were combined in favour of the higher commercial value of the forgery, and both were nearly lost to science. Palaeontology was also badly damaged by the Piltdown forgery¹⁰ and the 'lying stones' of Johann Beringer¹¹, and many fossils have been unwittingly or deliberately subjected to misleading reconstruction. Knowing the history of human

handling can be critical to proper evaluation and scientific interpretation of specimens. Fortunately, a growing array of techniques can now be applied to forensic analysis of fossils.

Timothy Rowe*, Richard A. Ketcham*, Cambria Denison*, Matthew Colbert*, Xing Xu†, Philip J. Currie‡

*High-Resolution X-ray Computed Tomography Facility, Department of Geological Sciences, C1100, The University of Texas at Austin, Austin, Texas 78712, USA

e-mail: rowe@mail.utexas.edu

†Institute of Vertebrate Paleontology and Paleoanthropology, Academia Sinica, PO Box 643, Beijing 100044, PR China

‡Royal Tyrrell Museum of Paleontology, Box 7500, Drumheller, Alberta T0J 0Y0, Canada

1. Sloan, C. P. *Natl Geogr. Mag.* **196**(5), 98–107 (1999).
2. Qiang, J., Currie, P. J., Norell, M. A. & Shu-An, J. *Nature* **393**, 753–761 (1998).
3. Xu, X., Wang, X.-L., Zhang, F.-C. & Wu, X.-C. *Nature* **401**, 262–265 (1999).
4. Dalton, R. *Nature* **403**, 689–690 (2000).
5. Dalton, R. *Nature* **404**, 696 (2000).
6. Holden, C. *Science* **288**, 238–239 (2000).
7. Rowe, T., Kappelman, J., Carlson, W. D., Ketcham, R. A. & Denison, C. *Geotimes* **42**, 23–27 (1997).
8. Dingus, L. & Rowe, T. *The Mistaken Extinction — Dinosaur Evolution and the Origin of Birds* (Freeman, New York, 1997).
9. Xu, X., Zhou, Z. & Wang, X. *Nature* **408**, 705–708 (2000).
10. Weiner, J. S. *The Piltdown Forgery* (Oxford University Press, London, 1955).
11. Jahn, M. E. & Woolf, D. J. *The Lying Stones of Dr. Johann Bartholomew Adam Beringer* (University of California Press, Berkeley, 1963).

Supplementary information is available on Nature's website at <http://www.nature.com> or as a paper copy from the London editorial office of Nature.

Liquid crystalline spinning of spider silk

Fritz Vollrath*† & David P. Knight*

* Zoology Department, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

† Zoology Department, Universitetsparken B135, 8000 Aarhus C, Denmark

Spider silk has outstanding mechanical properties despite being spun at close to ambient temperatures and pressures using water as the solvent. The spider achieves this feat of benign fibre processing by judiciously controlling the folding and crystallization of the main protein constituents, and by adding auxiliary compounds, to create a composite material of defined hierarchical structure. Because the 'spinning dope' (the material from which silk is spun) is liquid crystalline, spiders can draw it during extrusion into a hardened fibre using minimal forces. This process involves an unusual internal drawdown within the spider's spinneret that is not seen in industrial fibre processing, followed by a conventional external drawdown after the dope has left the spinneret. Successful copying of the spider's internal processing and precise control over protein folding, combined with knowledge of the gene sequences of its spinning dopes, could permit industrial production of silk-based fibres with unique properties under benign conditions.

Fabric fashioned of threads reeled from cocoons of the silkworm have long been one of the mainstays of world commerce, until the demands and constraints of the Second World War helped man-made fibres derived from petrochemical raw materials, or feedstocks, to supercede silk. Today, millions of tons of nylon are produced each year, selling for 2–10 US dollars per kilogram¹, whereas world silk production reaches only about 70,000 ton per year, with quality fibres selling for about 40 US dollars per kilogram. The current production rate and price of silk (<http://www.kanex.or.jp/groups/home/MARKET/XKitChart.htm>) are thus similar to that of top quality man-made fibres such as Kevlar and Twaron. However, silk polymers are poised for a possible comeback, albeit with a twist: artificial fibres spun from dope solutions composed of genetically modified (GM) natural proteins, designer proteins or protein–plastic blends^{2–5} could be the new 'techno-silks'.

Parent silk gene-motifs to create GM feedstocks for artificial dope solutions are likely to be taken from the golden silk spider *Nephila clavipes*, rather than from the moth *Bombyx mori*⁶, which has classically been the main silk provider. This preference is largely because spider silk is much tougher than the stiff but brittle cocoon threads of the moth, reflecting the fact that spider threads have evolved to arrest 'aerial missiles' rather than form a protective cradle. After all, silk has been an arachnid construction material for about 400 million years⁷, of which half has seen web-engineering responding to the demanding specifications of an aerial arms race with jumping and flying prey⁸. Accordingly, advanced web spiders have evolved an armory of highly specialized silks. (See Box 1b for an illustration of the range of mechanical properties of silks produced by different web spiders).

Spider dragline silks, the main structural web silk, which are also used for the spider's lifeline, are exceptionally strong and extensible⁶, making them very tough fibres. Indeed, their toughness equals that of commercial polyaramid (aromatic nylon) filaments, which themselves are benchmarks of modern polymer fibre technology. Polyaramids are used to make materials ranging from radial tyres and bulletproof clothing to reinforced composites for aircraft panels. The exceptional toughness of dragline silk is achieved under benign conditions: in contrast to current technopolymers based on petrochemicals, the spider spins its totally recyclable fibres at ambient temperatures, low pressures and with water as solvent^{9,10}. There are accordingly many advantages to copying the spider's silk and silk production capabilities¹¹: genetically engineered silk with specially tailored properties and spun using 'green' processes could replace the ubiquitous plastics, which are often detrimental to the environment in both production and

disposal³. In fact, spider silk has already been spun artificially: the most interesting methods until now include conventional solvent spinning of recombinant spider silk protein analogue by Du Pont Ltd² and extruding regenerated silk into a coagulation bath using a nanofabricated silicon spinneret¹². The race is now on to perfect spinning technology¹³. Of course, artificial silk could be a potential economic threat to rural producers of natural silks, but the overall benefits might still outweigh the costs and disadvantages, especially if the dope-feedstock used for the production of these techno-silks is derived from agricultural products, such as goat milk genetically modified to contain spidroins, the major proteins of spider dragline silk (see <http://www.nexiabiotech.com>).

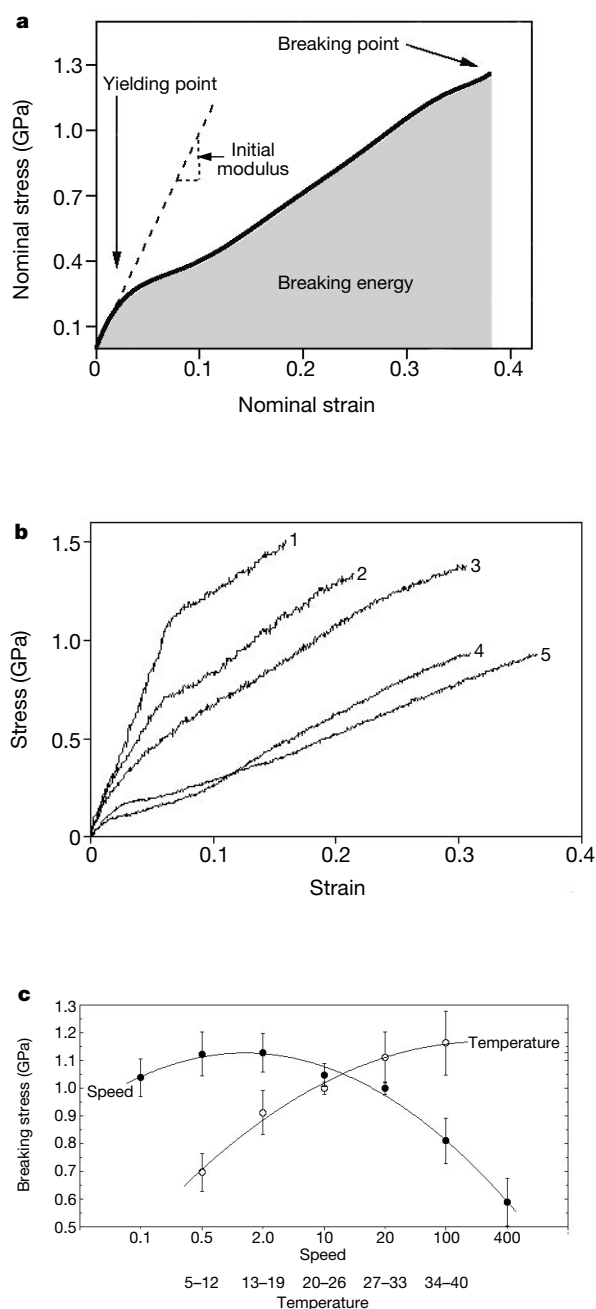
To mimic spider dragline silk successfully we must copy the crucial design features of both the feedstock dope proteins and the spinning process. Molecular biology provides the potential to extract, synthesize and assemble artificial genes to supply feedstocks for the production of silks^{4,14–16}, whereas classical morphological studies provide construction details of the spider's extrusion system^{17–21}. Together, these two biological disciplines can lead to the design of first prototypes for a highly advanced and benign fibre extrusion technology, and close collaboration with process engineering should eventually allow commercialization. Scientifically as well as technologically, achieving this goal would be an innovation. It would be the first time that a whole natural production system as complex as the spider's has been copied in totality from gene to marketable product, though there are many examples of industry copying an aspect of nature, including Velcro which mimics the way seed hooks snag in fur.

Composition of a thread

Spider and insect silks have the same basic building blocks: proteins made largely from non-essential amino acids^{22,23}. Although they evolved independently in a wide range of invertebrates²⁴, all but the faecal fibres of beetles are extrusions of specialist silk glands (Box 2). Insects have found many ways of making a wide range of silk types, but web spiders are unique in maintaining, in the same individual, a battery of highly specialized glands producing very different silks, often simultaneously²⁵ (Fig. 1). This diversity is driven by the central role played by silk in a spider's life and by the tight co-evolution of silk and behaviour such as prey capture, shelter construction and reproduction²⁶. Although the different silk glands in an advanced web spider like *Nephila* are likely to have evolved from a single type of gland, they now differ significantly from one another in morphology, anatomy and composition²⁷ to allow for the production of different silks (Fig. 1). Moreover, there seems to be significant variability in silk properties even for the same silk produced by

Box 1

Stronger than steel



Silk, like many man-made fibres, is viscoelastic⁷³; loading and unloading paths are not the same as they would be in a perfectly elastic material. Moreover, a small plastic component is also often observed (that is, the material does not recover fully to its original length upon relaxation), but this can be overcome by curing with water⁷⁴. The mechanical properties of a fibre are best described by the shape of a response curve when the material is stretched (**a**). Such stress-strain curves may vary greatly between the same type of silk from different spiders (**b**) and typically vary between the different silks from the same spider²⁶. The variability is due to differences in both the chemical composition of the dope and the conditions under which it is spun, once again emphasizing the importance of the spinning environment to control the material properties of the final silk thread (**c**). The spider

can easily vary spinning conditions by adjusting running speed, or changing speed of pulling silk with the legs (if used) and, finally, by building its webs during different times of the day or night (which affects temperature) all of which gives the animal some degree of control over the performance of the thread it produces.

Spider silk is light as well as strong. Its strength of 1.1 GPa approaches that of typical high tensile engineering steel (1.3 GPa), but silks have a significantly lower relative density (1.3) than steel (7.8) (ref. 73). Compared on a weight basis, silk is thus by far the stronger material. To probe the mechanical behaviour of silk (**c**), we stretch our silk at a specific rate dl/dt and measure the force F required to cause a given extension dl . We generally consider the engineering stress (s_E), which is the force applied to the cross-sectional area (A_0) of the sample, to be constant during the test, even though A_0 will decrease somewhat during stretching. The normalized engineering strain of extension (e_E) is the ratio of change in length (dl) to the initial gauge length (l_0). The ratio of stress to strain, given as Young's modulus (E) and derived from the slope of the curve, is a measure of the fibre's stiffness. Typically there is at least one 'yield point', that is, a point at which the curve suddenly changes slope because of major structural transitions in the material. The area under the curve (if the fibre breaks) or between hysteresis curves (if the fibre is relaxed before it breaks) indicates the fibre's toughness, and thus the energy taken up by the material (**a**). **a**, Typical stress-strain curve for spider dragline silk reeled from a *Nephila edulis* female, based on experimental data. The data are obtained from stretching experiments on dragline silk, giving the force (stress) required to reach a given strain, with strain indicating the fibre's extension under stress relative to its extension in the relaxed state. Average data (mean $\pm$ standard deviation, s.d.) for silk collected at 20 mm s^{-1} and 25°C from an adult *Nephila edulis* female are: silk diameter, $3.35 \pm 0.63 \mu\text{m}$; tensile breaking strain, 0.39 ± 0.08 ; breaking stress, $1.15 \pm 0.20 \text{ GPa}$; initial modulus, $7.9 \pm 1.8 \text{ GPa}$; yield stress, $0.15 \pm 0.06 \text{ GPa}$; and breaking energy, $165 \pm 30 \text{ kJ kg}^{-1}$ (after ref. 63). Comparable data for Kevlar 81 high tenacity yarn 98 are: diameter, $12 \mu\text{m}$; breaking strain 0.05; breaking stress 3.6 GPa; modulus 90 GPa; no yield stress as yarn has a single modulus; and breaking energy around 33 kJ kg^{-1} (A. Grandus, personal communication). Thus Kevlar is 3 times stronger but spider silk is 5 times tougher because it is 8 times more extendible. **b**, Stress-strain characteristics of dragline silk reeled from different web building spiders: *Euprosthenoops* sp (Pisauridae) (**1**), *Cyrtophora citricola* (Araneidae) (**2**), *Latrodectus mactans* (Theridiidae) (**3**), *Araneus diadematus* (Araneidae) (**4**) and *Nephila edulis* (Tetragnathidae) (**5**). The data (modified from ref. 28) were collected under comparable conditions by artificial reeling. They show that the silk of *Euprosthenoops* (**1**) is stiffer and requires more force to break it but is much less elastic and thus takes up less energy than the comparable silk of *Nephila* (**5**); the characteristics of the other species lie in between. We note also the differences in the initial moduli and yielding points. **c**, The effect of spinning conditions on the breaking stress of dragline silk spun by *Nephila edulis*. The black circles give the effects of modifying reeling speed (in mm s^{-1}) and the white circles those of modifying body temperature (in $^\circ\text{C}$) (figure from ref. 63). The average breaking energy for the sample threads at control conditions was $165 \pm 30 \text{ kJ kg}^{-1}$ (ref. 63). In the graph, breaking stress is used rather than breaking force because the differences between individual spiders (as well as silk reeling speed and body temperature) affect the diameter of the silk thread, which in turn influences the force needed to break the thread (see ref. 63). The control temperature was 25°C when the reeling speed was varied, and the control speed was set to 20 mm s^{-1} when the body temperature was varied⁶³, with the average natural spinning conditions for this species taken to match the control conditions⁶³. (Each data point derives from threads of 4–11 spiders with four pieces of each thread measured, mean $\pm$ s.d.; figure modified from ref. 63.)

the same spider on different days, or even on the same day under different environmental conditions²⁸. Most of these differences are attributable to varying spinning conditions, but they may also reflect considerable variability in the amino acid composition of the spinning dope produced by a gland^{22,23,29}, some of which might be affected by diet^{29,30}.

Gene sequences are being mapped for rapidly increasing numbers of silks^{15,16,31–34} and first-generation models are beginning to relate protein structure to the mechanical properties of the thread^{35,36} (see Box 3). But although gene sequences can predict average domain sizes along the protein backbone, any deeper understanding of the mechanical properties of spider silk requires additional information on the interaction between the different proteins in a thread and the full structure at all hierarchical levels. And even though spider silk seems a poor cousin in terms of diversity, with only two or three spider silk protein gene families identified so far^{15,16,33} (a large number of different proteins, corresponding to many genes, have been identified in the silks of some chironomid midges³⁷), shunting and splicing would allow a spider to create a wealth of spider silk proteins from quite a parsimonious genetic blueprint^{29–31}. More-

over, expression differences in each of the spider's seven different gland types allow transcriptional regulation to produce a wide variety of gland-specific silks with different compositions and material properties¹⁶.

At least one spider silk protein family (found, for example, in the golden silk spider *Nephila clavipes* and the garden cross-spider *Araneus diadematus*) encodes for polypeptides that contain a variable number of both crystalline poly-alanine domains as well as less-crystalline glycine-rich domains^{15,16,32,37}. Although there is consensus that silk contains crystalline and non-crystalline regions, the structure of the glycine-rich regions is imperfectly understood. Indeed, the classic concept of dragline silk being comprised of β -sheet crystals packed in a rubbery matrix of random-coiled springs^{35,36,38} has been challenged^{39,40} following the observation that silk may contain a range of conformations such as 'random coil', β -turns and parallel and/or antiparallel β -pleated sheet crystals, α -helices or the more compact and left-handed 3(1)-helices^{35,36,38–40} (see Box 3 for further details). Clearly, before we can copy silk to full effect we must determine the secondary structure of polypeptide chain folding as well as higher-order structures of the protein constituents. For example, recent studies have shown that a single spider silk filament can have a complex 'coat-skin-core' organization^{41,42} with micro-fibrils^{43–45} enclosing fine channels⁴¹ that might impart energy-dispersive properties⁴⁶ by deflecting the tips of cracks forcing their way across the thread. Because this structural complexity is likely to contribute significantly to the toughness of spider silk, it must be taken into account when attempting to design and spin biomimetic silks.

Box 2

Other spinning arthropods

Bombyx mori is a domesticated moth with a larva that feeds on mulberry leaves. Its pupation cocoon is unravelled to give the classic commercial silk; other commercial silks come from other moths, such as *Antheraea*. Our knowledge of the spinning processes in these insects is extensive but incomplete^{48,75}. The paired silk glands are long and heavily tracheated ectodermal ducts that open on the lips of the mouth. Silk fibroin and droplets, which later form canaliculi, are secreted at the posterior end of the gland and move along in laminar flow while being coated by three layers of sericin, the silk's second protein constituent^{76,77}. The first, thin coating contains one polypeptide (s-4), the second, thicker granular layer contains two (s-1 and s-3) and the thin, third layer contains a further two polypeptides (s-2 and s-5)^{48,77}. All these polypeptides are rich in glycine, serine and aspartic acid, and s-1 and s-2 also contain sugars⁷⁷. Thus coated, the silk filament is drawn (the larva pulls by moving its head) through a cuticle-lined duct to join its companion filament made in the other branch of the bilaterally symmetrical silk organ. Both filaments are then glued together by secretions from Filippi's gland, with the still pliable sericin being moulded by the silk press (a muscular valve) and the labial opening^{75,77}. A silkworm silk gland thus consists of a posterior part producing silk-dope, three distinct middle parts producing sericin-coat, a tapering duct, a Y-junction joining the two glands, followed by a single press and opening. In a cocooning *Bombyx* larva, this organ constitutes about 20% of the animal's body weight, and a single cocoon will yield about 800 m of silk fibre composed of two filaments, each with an average diameter of about 10 μ m (ref. 75). The heavily coated fibres are stiff and strong, but not very tough because they lack extensibility. Silk production is not confined to moths and spiders²⁴: bees and wasps (Hymenoptera) embed simple silk fibres into the wax of their combs to provide strength⁷⁸, fleas (Siphonaptera) use silk in their nests and cocoons⁷⁹, lacewings (Neuroptera) employ it to cement eggs onto stalks⁸⁰, caddisfly larvae (Trichoptera) use it to bind debris into houses or make underwater webs⁸¹, aquatic midge larvae (Diptera) use it to construct feeding nets and tubes for housing and pupation³⁷ and glow worms (Coleoptera) and fungus gnats (Nematocera) make sticky silk lines to catch flies⁸². Among the arachnids, silk is used for nest building in the pseudoscorpions and mites⁸³, and in all aspects of life in the true spiders (Araneids)¹⁹. Diversity in taxonomy, modes of deployment and evolutionary pathways resulted in a wide range of molecular chemistries, material properties and spinning technologies²⁴ associated with silk production, thus providing ample scope for biomimetic material design and processing.

Spinning a fibre

Let us take the production of dragline silk by *Nephila* as an example of a spider's advanced spinning technology, given that the process is best understood in this species and probably representative for dragline spinning by all orb-spiders. Figure 2 gives a schematic illustration of the different components of the spinning duct, the structure through which the spider extrudes its silk and which is analogous to the 'die' used in artificial fibre spinning technology. The major ampullate gland making this silk consists of a long tail and a wider sac. The tail secretes the major part of the so-called 'spinning dope', a solution containing the protein molecules used to make the silk fibre dissolved in water, while the sac constitutes the main storage repository that leads, through a 'funnel', to a tapering duct. Much of the fibre formation occurs in this duct, which has three loops inside a sheath and terminates in a structure known as the valve. After the valve, further processing proceeds in a narrow tubular region specialized for rapid water recovery, and the silk thread then exits at the spigot whose flexible lips fit tightly round the silk thread, a feature assumed to assist in water retention²¹. Tail and sac of the gland are about 60 mm in length and, when full, weigh in at about 10 mg the pair, about 1% of the total bodyweight of a lean adult female (D.P.K. and F.V., unpublished work).

The secretory part of the gland has two distinct transverse zones, the A-zone occupying the tail and first part of the sac, and the B-zone running from the widest part of the sac to the funnel. The epithelium of the A-zone is composed of tall columnar secretory cells of a single type, packed with secretory granules (see Fig. 2, inset A). The output of these cells is an aqueous and highly viscous, often yellow solution of about 50% protein⁴⁷, which will be mostly spidroin I and II, the main proteins making up spider dragline silk². The solution secreted by the A-zone is not a homogeneous liquid, but contains many small spherical droplets comparable to those found in silkworm silk dope⁴⁸. These droplets flow along the tail and sac⁴⁹ in the spidroin matrix, and are thought to first grow by coalescence before being drawn out into long thin structures, the canaliculi^{41,49}. As the A-zone secretion flows towards the funnel it is coated by a colourless homogeneous viscous liquid secreted in the B-zone. Although the cells of the B-zone superficially resemble

those of the A-zone (see Fig. 2, inset B), they contain different secretory granules, many of them filled with a hexagonal columnar liquid crystalline material. Given the relative size and locations of these zones, we may assume that the more extensive A-zone (about 75% of the epithelium) secretes the spidroin(s), whereas the smaller B-zone in orb-web spiders secretes coat protein(s), possibly the glycoprotein identified histochemically in the gland¹⁷ and biochemically in the thread⁵⁰. The glycoprotein may help to plasticize the thread by maintaining a high water content⁵¹.

Within the major ampullate gland and the first and second loop, or limb, of the spinning duct, the spider's dope^{49,52,53}, like the silkworm's^{54,55}, is liquid crystalline, with the main silk protein constituent likely to be in a compact conformation that allows it to be processed at high concentrations⁴⁷. Liquid crystallinity offers desirable properties, which make it possible to efficiently spin a thread from molecules as large as silk proteins (see Box 4). Specifically, in the spider's gland and duct the molecules seem to form a nematic phase⁴⁹, that is, they form a substance that flows as a liquid but maintains some of the orientational order characteristic of a crystal, with the long axes of neighbouring molecules aligned

approximately parallel to one another. Liquid crystallinity would thus allow the viscous silk protein solution to flow slowly through the storage sac and duct while the molecules form complex alignment patterns. In the spider's gland, the long axes of the rod-shaped silk protein molecules or molecular aggregates appear⁴⁹ to be oriented perpendicular to the secreting epithelium walls when close to them, but gradually bend over with increasing distance until—along the midline of the glandular sac—they lie parallel to the long axis. This arrangement (called a simple escaped nematic texture) persists from the gland into the first limb of the duct, but changes into a more complex escaped nematic texture in the second limb of the duct, where the narrower lumen forces the molecules to bend forwards and backwards in an alternating pattern known as the cellular optical texture^{49,56}. This arrangement is commonly seen when 'nematic discotic' liquid crystals are confined in narrow tubes⁵⁶. This type of liquid crystal forms bilayered disks in which the rod-shaped molecules are arranged perpendicular to the plane of the disk⁵⁶. It seems likely that the perpendicular arrangement of the silk protein molecules at the secreting epithelium walls and the subsequent nematic escape into an arrangement that is parallel to

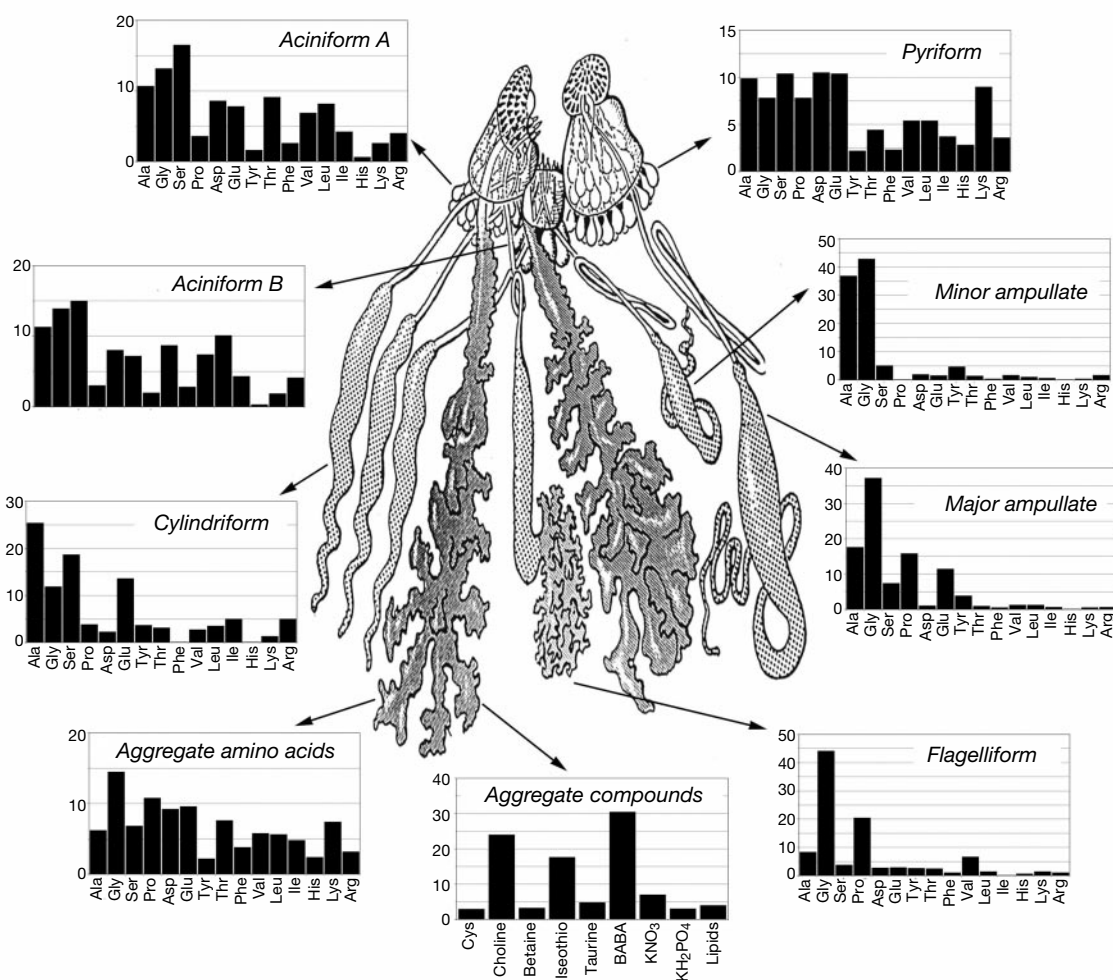


Figure 1 The seven specialized glands and their different amino acid compositions of a typical Araneid orb weaver. Silk glands are paired, reflecting their evolution from limb glands⁶⁴. The aciniform glands make the silk for wrapping prey, the cylindriform glands make the cover silk for the egg sac and the pyriform glands make the silk for attachments and for joining threads^{25,27}. The major and minor ampullate glands are of similar shape; the major glands provide the silk for the spider's dragline and the frame threads of its web while the minor ones provide threads that can be added to any structural thread and may be used for the radial (spoke) threads of the web^{25,27}. The flagelliform glands provide the core of the threads of the capture spiral, while the threads' coating is supplied by the aggregate glands. This coating contains glycoproteins for stickiness⁶⁵, salts to act as bactericides⁶⁶ and hygroscopic organic compounds⁶⁷ that, in addition to being bactericidal, also attract atmospheric water⁶⁸ to plasticize the core of the thread⁶⁹. (Glands after ref. 70 and amino acid compositions after ref. 22; we note there can be variability in composition within each gland, which may be adaptive²⁹).

the long axis of the silk gland and spinning duct prevent the liquid crystalline dope from breaking up into numerous small domains. This in turn would suppress the formation of disclinations, a form of defect somewhat analogous to dislocations in solid crystals and which are likely to diminish the tensile strength of the spun thread^{49,57}.

The duct's convergent, or hyperbolic, geometry forces the dope flowing along it to elongate at a practically constant rate⁴⁹. As a result, the spherical droplets in the dope extend to form the long, thin and axially orientated canaliculi^{21,49}, which are thought to contribute to the thread's toughness⁴⁶. The slow, constant nature of the elongation also ensures that only low and uniform stresses are generated. This prevents localized coagulation centres from forming prematurely⁴⁹, that is before the silk protein molecules in the dope have reached their optimal orientation. As in any other spinning, good molecular alignment contributes significantly to the thread's toughness^{2,3,57,58}.

A much higher stress is likely to be generated during rapid extension in what we have called the "internal drawdown taper"²¹ (see Fig. 2, panel X). This starts where the forming thread suddenly stretches, narrows and pulls away from the walls of the third limb of the duct (in the fully grown spider approximately 4 mm before the exit spigot)^{21,49} (see below). The high stress forces generated during this stage of processing probably bring the dope molecules into alignment and into a more extended conformation, so that they are able to join together with hydrogen bonds like zip fasteners to give the anti-parallel beta conformation of the final thread. As the silk protein molecules aggregate and crystallize, they will become more hydrophobic, which should induce phase separation and hence the loss of water from the surface of the solidifying thread. Recent studies of the flow behaviour of dope solutions *in vitro* (Chen, X.,

D.P.K. and F.V., unpublished work) and the observation that the force needed to reel silk *in vivo* increases with reeling speed⁵⁹ (except at very high reeling speeds) provides direct evidence that such a stress-induced phase separation does indeed occur. The secretion of hydrogen ions, by proton pumps located in the epithelial cells of the third limb of the duct, renders the dope solution more acidic, which should assist phase separation^{20,59}. The concluding drawdown of the silk, to produce the final fibre, occurs in the air after the thread has left the spigot^{21,60} and is probably accompanied by further loss of water by evaporation^{53,60}. However, as some natural silks are successfully spun underwater³⁷, water loss by evaporation is not always necessary.

The spider's simultaneous and internal drawdown and material processing which exploits phase-separation differs dramatically from that used in industrial spinning, where the solvent escapes in a gradient or to the surface, typically at the die's external opening⁵⁷. Although the latter type of processing, with external drawdown into air, may occur in the spinning glands of other spiders⁶¹, keeping the drawing process inside the animal appears to offer conspicuous advantages: most of the water from the dope can be recycled by absorption from the distal part of the duct⁶², and the duct acts as a combined internal die and a treatment and coating bath²⁰.

As the fibre forms from the dope, it pulls away from the surrounding walls of the spinneret. This initial phase of the draw-down should start at that location in the duct where the force required for drawing out a thread by extensional flow balances or just falls beneath the force required for moving along the liquid dope in contact with the walls. As it moves through the spinneret, the forming thread is probably mostly lubricated by water formed in the phase separation. A surfactant secreted into the third limb of the

Box 3

Silk proteins and genes

Spider silk proteins are only partially characterized, even for *Nephila clavipes*⁶⁴, the spider species that has been most studied and which is thus our benchmark. Separating the proteins by sodium dodecyl sulphate–polyacrylamide gel electrophoresis suggests that the major ampullate gland in this species secretes a single major protein component, whose apparent molecular weight has been determined as 275 kDa (ref. 85) and 320 kDa (ref. 86). The parent protein is about 5% glycosylated⁶⁰ and seems to contain sequences derived from Spidroin I and one or more additional domains not seen in either Spidroin I or II (ref. 85). Initial characterizations suggest that the protein is constructed from two peptide chains held together by 3–5 disulphide bonds⁸⁵, which may aggregate under native conditions to give an oligomer with a molecular weight of about 725 kDa (ref. 2). Another component, with an estimated molecular weight of 180 kDa, has been identified in mercaptoethanol-reduced silk⁸⁵. Until now, only a few silk genes have been partially sequenced, and most of these are from orb spiders⁶⁴. Spidroin I and Spidroin II are two subunits of dragline silk found in the major ampullate glands⁶⁴. A very different protein, containing only three motifs repeated many times in the same order, is secreted by the flagelliform glands³¹; it forms a major component of the elastic support fibre of the capture spiral of the spider's web. The largest of the protein's three motifs, (Gly-Pro-Gly-X)_n, may form a β-spiral or helix composed of a large number of β-turns arranged end-to-end; these can then act as springs, which bond together via water bridges at the X-residues to form a highly elastic network⁸¹. In contrast, the sequence data for Spidroin I and II show an alternation of hydrophobic polyaniline domains and less hydrophobic glycine-rich domains. If we are to understand the way in which the protein refolds during spinning, we need to clarify its structure during production and storage in the gland cells and sac. Spidroin may form rod-shaped molecular or supramolecular units, which can be stored as a highly concentrated liquid crystalline solution to form a fibre when

strained in the spinning duct⁴⁹. It seems that unspun spidroin differs in conformation from the silk thread, but resembles silk-I of silkworm dope⁴⁷. NMR and ultraviolet circular dichroism⁴⁷ measurements on unspun spider dope indicate that it lacks the β-pleat seen in the final silk and instead contains up to about 30% α-helices, 40% of a 'random coil' conformation, and 30% β-turns that might exist in a loose and dynamic, partially extended configuration analogous to that of the molten globule state. Taken together, the observational evidence suggests a multiple β-turn structure for the unspun dope; this might contain triple⁸⁷ or double⁸⁸ β-turns stabilized by interstrand hydrogen bonds, or a β-turn/'random coil'/α-helical structure⁴⁷. The location of the β-turns and α-helix motifs is not known with certainty, but in aqueous solution, the glycine-rich regions and the polyaniline sequence are likely to form the β-turns and α-helices, respectively⁸⁴. Advanced imaging techniques such as NMR, micro-Raman spectroscopy and micro-X-ray diffraction are crucial for analysing the structure of the secondary and tertiary conformations of the silk dope and thread. For example, proton-driven ¹³C two-dimensional NMR spin-diffusion experiments on *Nephila* dragline silk thread suggest that the poly-Ala segments adopt a highly ordered β-sheet structure^{39,47} and the glycine-rich segments 3(1) helical structures^{40,88}, both oriented in line with the thread. Raman spectroscopy on single *Bombyx* and *Nephila* threads during progressive stressing shows that the silk microstructure experiences uniform stress during deformation in both spider and silkworm silk^{89–91}, suggesting the presence of microfibrillar structures inside the thread that would evenly distribute any applied stress. X-ray diffraction patterns obtained from single threads being spun⁹² also reveal that the thread consists of small crystalline blocks in a matrix containing both oriented and disoriented amorphous material, with the crystalline fraction having a β-poly (L-alanine) structure⁹³.

duct may also help to reduce the force required for drawdown. The tapered nature of the duct may allow the point of drawdown to move along the duct according to the rate of spinning, thus resulting in thread diameters adjusted to the speed of spinning⁶³.

In summary, spiders make exquisite use of internal spinning of liquid crystalline dope in an acid bath, followed by further drawdown in the external air gap, to produce silk threads of high toughness and elasticity. Spinning of liquid crystalline material, as

done by spiders, has several advantages: (1) it is virtually free of the uncontrolled reorientation of molecules after emergence from the die; (2) the force required to spin a thread is small; and (3) pre-alignment of the molecules in the unspun dope may reduce the formation of defects⁵⁷. It seems that much stands to be gained by trying to understand and copy the spinning technology, which the spider has evolved over millions of years to regulate the folding and refolding of macromolecules under highly controlled conditions to

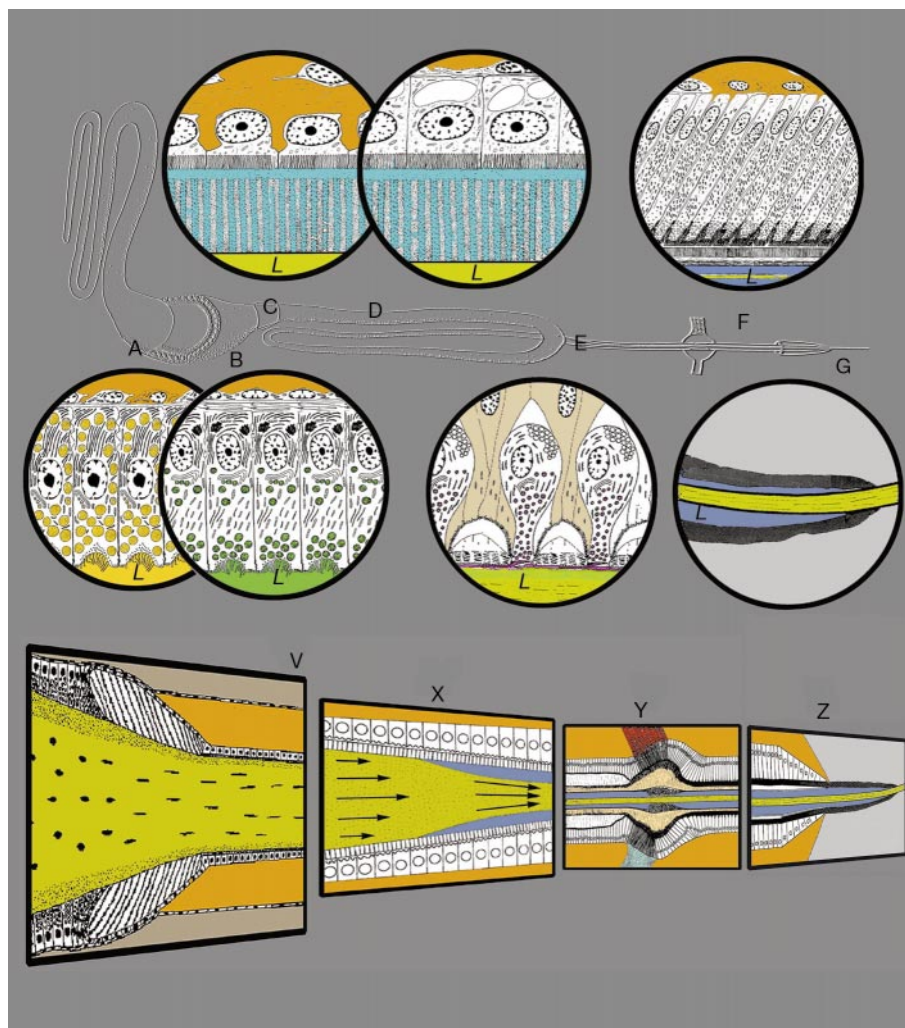


Figure 2 A spider's dragline spinneret. The top part shows original drawings of the histology of the spinneret (lumen, L); the bottom part outlines its spinning function, which entails drawing the liquid crystalline dope solution produced in the gland through a tapering s-shaped duct, thereby converting it into an elastic thread. Dope production occurs in two zones: the A-zone of the gland (see panel A) secretes spidroin, the protein forming the core of the thread, while the B-zone (see panel B) secretes the thread's thin coat. The secretory vesicles of the A-zone contain short, narrow filaments; most of those of the B-zone contain hexagonal columnar liquid crystals. The greatly thickened cuticle of the funnel (clearly visible in panel V) anchors the extensible duct to the less mobile sac, perhaps to prevent shearing of the dope when the spider wiggles. The duct itself has a thin cuticle, which acts as a dialysis membrane and may allow water²⁰ and sodium ions out of the lumen, and potassium ions⁷¹, surfactants and lubricants into the lumen to facilitate thread formation (see panel X). The epithelium of the s-shaped duct progressively increases in height from the funnel to the valve, suggesting increased pumping of water and ions as the dope is drawn through the duct (panels C and D show the histology of epithelium in the first and second limbs, respectively). The drawdown process is mainly internal and starts in the third limb of the duct (see panel X), approximately 4 mm from the exit spigot. Just before and after the start of the drawdown taper, single flask-shaped gland cells, such as seen in inset E, may contribute an extra coating to the thread²¹. The 'valve', shown in panel Y in the bottom of the figure, is not a restriction nozzle, but a clamp for gripping the thread²¹. It may also act as a 'ratchet' or 'pump', to restart spinning after internal rupture of a filament²¹. The section of duct (see inset F) after the valve appears to be highly specialized for water pumping, having tall cells with numerous mitochondria and apical microvilli and a large surface area provided by apical infoldings of the plasma membrane²⁰. Finally, the thread is gripped by the flexible and elastic lips of the spigot (inset G), through which it passes to the outside world (see panel Z). The spigot strips off the last of the aqueous phase surrounding the thread, thus helping to retain water in the spider, and also places the thread under tension for the final air-drawing step^{21,60}. The diameters of the round panels are A: 93 μm , B 93 μm , C 12 μm , D 12 μm , E 70 μm , F 288 μm , G 60 μm (drawings based on light microscopy, SEM, CryoSEM energy dispersive X-ray analysis and TEM studies of the major ampullate gland of three species of *Nephila*); the heights at midpoint of the frame panels are V (funnel) 350 μm , X (drawdown) 40 μm , Y (valve) 300 μm , Z (spigot) 190 μm .

Box 4

Liquid crystalline spinning

The rheology of liquid crystalline solutions is non-newtonian and differs from that of isotropic solutions, a feature that can be exploited by a biological system under intense pressure to adapt and optimize. In particular, while liquid crystalline viscosity depends on the shear rate, stress forces along and across the flow lines of the liquid are highly dependent on the concentration of the solution and can be very small indeed^{57,58}. By controlling the solvent saturation (that is, the water content) of the liquid crystalline spinning solution, the spider can thus vastly reduce its metabolic cost of silk production and use narrow extrusion ducts — which would otherwise require high energy input to operate — to produce very fine fibres ranging from a few nanometres to a few micrometres in diameter^{57,94}. Newtonian laminar flow (where the shear rate determines the stress forces but not the viscosity), in contrast, is characterized by a nonlinear relationship between the coefficient of flow viscosity, h_t , and the tube radius, R , (that is, $h_t = pR^4/8Q \cdot dp/dl \cdot \eta_1$), as well as the speed of spinning (through Q , the volume rate of flow, and dp/dl , the gradient of the pressure p along the tube length l)⁹⁴. Clearly, making fibres under newtonian conditions would require very high pulling forces that could not be generated by a spider walking²⁵ either at normal speed (20 mm s⁻¹) or rushing at 800 mm s⁻¹. Moreover, in newtonian flow, viscosity increases with decreasing temperature: $h_t = h_0 \exp(E_a/RT)$, where h_0 is the basic viscosity and E_a the activation energy, which depends on the molecular structure of the polymer⁹⁵. The relative temperature independence of non-newtonian flow, in contrast, would benefit a spider by allowing it to spin webs in the cold early hours of the morning as easily as in the heat of noon. Both spider and moth have found ways to employ non-newtonian fluid dynamics in order to make thin silk filaments at ambient temperatures and low energy costs, by using liquid crystallinity, well adjusted protein concentrations and reasonably low spinning speeds (and hence shear rates). Shear thinning, a decrease in viscosity resulting from moderate increases in shear rate has been observed in nematic liquid polymers and also in spider silk dope solutions (Chen, X., D.P.K. and F.V., unpublished work); it will further help to reduce the energy required to flow the dope through the spider's spinning duct. Although liquid crystalline polymers behave as newtonian fluids at high dilutions, at high concentrations their flow is non-newtonian (that is, viscosity depends on strain rate and strain history) and the apparent viscosity at first increases with concentration but then actually decreases to initially low levels⁵⁸. This is because the molecules tend to align at high concentrations, enabling them to slide or roll past each other more easily⁵⁷. Rod-like liquid crystalline molecules with length L in solutions (of concentrations c) diluted to $c \ll 1/L^3$ have a viscosity

$h = h_s(1 + cL^3)$, which is practically that of the solvent; in solutions where $c \gg 1/L^3$, the viscosity greatly increases to $h = h_s(1 + (cL^3)^3)$, and is thus much more dependent on molecular length and therefore on molecular weight⁵⁷. Liquid crystalline viscosity is strongly affected by the vibrational freedom of the molecules, making it very sensitive to both the rate and history of shear and thus leading to strong non-newtonian viscoelasticity⁵⁷. Spinning from lyotropic solutions (solutions that are rendered liquid crystalline by adding solvent molecules to liquid crystalline polymer) allows for high axial alignment of the liquid crystal molecules and thus affects the spinning performance, particularly if, as in the spider, there is a long region over which the fibre is drawn out by progressively increasing the linear flow rate of the dope. This also ensures molecular orientation throughout the fibre before the molecules are locked into position by a coagulation bath (in wet spinning) or in an air gap (in dry spinning). In fact, in order to realize the outstanding elastomeric properties of spider silk, the fibre's molecular rods and springs need to be well aligned; high-tenacity fibres such as Kevlar have, in contrast, very few in-built 'molecular springs' analogous to the spider's helices or β -turns, which explains the low elasticity of Kevlar. The absolute size of the molecules and their size distribution⁵⁷ are also important parameters affecting the toughness of the final thread: large variability in molecular size leads to incomplete molecular bonding. The toughness of spider dragline silk may thus depend in part on the fact that it contains predominantly a single large protein^{84,85} with little variability in molecular weight², achieved by the tight control inherent in nature's way of making proteins. Furthermore, the large molecular weight of this protein (~ 250 kDa) implies that the silk protein molecules cannot be fully extended in the dope; if they were, the material would turn into a gel at the high protein concentrations found in the storage sac and duct. In fact, molecules with a molecular weight over 30 kDa are difficult to spin industrially⁹⁶ because polymers with high molecular weight produce highly viscous spinning dopes. For example, the terephthalamides used for the production of DuPont's Kevlar and Acordis's Twaron have molecular weights of only 13–40 kDa, which is an order of magnitude smaller than that of spider silk proteins. These para-aramids, para-oriented aromatic nylons, are spun from concentrated solutions in concentrated sulphuric acid through a series of water baths, using highly controlled and narrow ranges of spinning rates and temperatures⁹⁵. Clearly, spiders produce high quality fibres using a different method: they optimize dope viscosities and molecular conformations as well as spinneret design while keeping spinning conditions much more flexible⁶³.

allow their efficient extrusion into a thread. An important lesson to learn from the spider is how it stores protein dope molecules in a highly concentrated liquid crystalline state and then extends these in the spinning duct to form a supremely tough thread. Learning this lesson may also help us to gain insight, through lateral thinking, into the sudden but unwanted assembly of other fibrous proteins, such as amyloids (the protein associated with Alzheimer's disease) that refold and form filaments where we had rather they did not. $\square$

1. Fiber Economics Bureau. *Manufactured Fiber Review 2000*, Jan 2–53 (Fiber Economics Bureau, Washington, 1999).
2. O'Brien, J. P., Fahnestock, S. R., Termonia, Y. & Gardner, K. C. H. Nylons from nature: synthetic analogs to spider silk. *Adv. Mater.* **10**, 1185–1197 (1998).
3. Tirrell, J. G., Fournier, M. J., Mason, T. L. & Tirrell, D. A. Biomolecular materials. *Chem. Eng. News* **72**, 40–51 (1994).
4. Prince, J., McGrath, K., Digirolamo, C. & Kaplan, D. Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry* **34**, 10879–10885 (1995).
5. Sun, Y., Shao, Z., Hu, P. & Tu, T. Hydrogen bonds in silk fibroin-poly(acrylonitrile-co-methyl acrylate) blends: FT-IR study. *J. Polym. Sci. B* **35**, 1405–1414 (1997).
6. Kaplan, D. L., Adams, W. W., Viney, C. & Farmer, B. L. (eds) *Silk Polymers: Materials Science and Biotechnology* 1–370 (ACS Books, Washington, 1994).
7. Shear, W. A., Palmer, J. M., Coddington, J. A. & Bonamo, P. M. A Devonian spinneret: early evidence of spiders and silk use. *Science* **246**, 479–481 (1989).

8. Selden, P. A. Orb-weaving spiders in the early Cretaceous. *Nature* **340**, 711 (1989).
9. Heslot, H. Artificial fibrous proteins: a review. *Biochimie* **80**, 19–31 (1998).
10. Asakura, T. & Kaplan, D. L. Silk production and processing. *Encyclop. Agric. Sci.* **4**, 1–11 (1994).
11. Kaplan, D., Adams, W. W., Farmer, B. & Viney, C. in *Silk Polymers. Materials Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 2–16 (American Chemical Society, Washington, 1994).
12. Seidel, A., Liivak, O. & Jelinski, L. W. Artificial spinning of spider silk. *Macromolecules* **31**, 6733–6736 (1998).
13. Vollrath, F. & Knight, D. P. Apparatus and method for forming materials. PCT Patent Application PCT/GB00/04489 (1999).
14. Kaplan, D. L. & Lombardi, S. J. The amino acid composition of major ampullate gland silk (Dragline) of *Nephila clavipes* (Araneae, Tetragnathidae). *J. Arachnol.* **18**, 297–306 (1990).
15. Winkler, S. & Kaplan, D. L. Molecular biology of spider silk. *Rev. Mol. Biotech.* **74**, 85–93 (2000).
16. Guerette, P., Ginzing, D., Weber, B. & Gosline, J. Silk properties of determined by gland-specific expression of a spider fibroin gene family. *Science* **272**, 112–115 (1996).
17. Kovoov, J. La soie et les glandes sericigenes des arachnides. *Ann. Biol.* **16**, 97–171 (1977).
18. Kovoov, J. & Zylberberg, L. Fine structural aspects of silk secretion in a spider. *Tissue Cell* **14**, 519–530 (1982).
19. Kovoov, J. in *Ecophysiology of Spiders* (ed. Nentwig, W.) 160–186 (Springer, Berlin/Heidelberg/New York, 1987).
20. Vollrath, F., Wen Hu, X. & Knight, D. P. Silk production in a spider involves acid bath treatment. *Proc. R. Soc. B* **263**, 817–820 (1998).
21. Vollrath, F. & Knight, D. P. Structure and function of the silk production pathway in the spider *Nephila edulis*. *Int. J. Biol. Macromol.* **24**, 243–249 (1998).
22. Andersen, S. O. Amino acid composition of spider silks. *Comp. Biochem. Physiol.* **35**, 705–711 (1970).
23. Work, R. W. & Young, C. T. The amino acid compositions of major and minor ampullate silks of

- certain orb-web-building spiders (Araneae, Araneidae). *J. Arachnol.* **15**, 65–80 (1987).
24. Craig, C. L. Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**, 231–267 (1997).
 25. Foelix, R. *Biology of Spiders* (Oxford Univ. Press, Oxford, 1996).
 26. Vollrath, F. in *Biomechanics in Animal Behaviour* (eds Domenici, P. & Blake, R. W.) 315–334 (Bios, Oxford, 2000).
 27. Vollrath, F. Spider webs and silk. *Sci. Am.* **266**, 70–76 (1992).
 28. Madsen, B., Shao, Z. & Vollrath, F. Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int. J. Biol. Macromol.* **24**, 301–306 (1999).
 29. Vollrath, F. Biology of spider silk. *Int. J. Biol. Macromol.* **24**, 81–88 (1999).
 30. Craig, C. L., Hsu, M., Kaplan, D. & Pierce, N. E. A comparison of the composition of silk proteins produced by spiders and insects. *Int. J. Biol. Macromol.* **24**, 109–118 (1999).
 31. Hayashi, C. Y. & Lewis, R. V. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Science* **287**, 1477–1479 (2000).
 32. Hinman, M., Dong, Z., Xu, M. & Lewis, R. W. Spider silk: a mystery starting to unravel. *Mater. Res. Soc. Symp.* **292**, 25–34 (1993).
 33. Hayashi, C. Y., Shipley, N. H. & Lewis, R. V. Hypotheses that correlate the sequence, structure and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **24**, 271–275 (1999).
 34. Kaplan, D. L., Prince, J. T., McGrath, K. P. & Digirolamo, C. M. Construction, cloning and expression of synthetic genes encoding spider dragline silk. *Biochemistry* **34**, 10879–10885 (1995).
 35. Gosline, J. M. et al. in *Silk Polymers. Materials Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 328–341 (American Chemical Society, Washington, 1994).
 36. Termonia, Y. Molecular modeling of spider silk elasticity. *Macromolecules* **27**, 7378–7381 (1994).
 37. Case, S. T. & Thornton, J. R. High molecular mass complexes of aquatic silk proteins. *Int. J. Biol. Macromol.* **24**, 89–101 (1999).
 38. Gosline, J., Denny, M. & DeMont, M. Spider silk as rubber. *Nature* **309**, 551–552 (1984).
 39. Simmons, A., Michal, C. & Jelinski, L. Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84–87 (1996).
 40. Kümmerlen, J., van Beek, J., Vollrath, F. & Meier, B. Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules* **29**, 2920–2928 (1996).
 41. Frische, S., Maunsbach, A. B. & Vollrath, F. Elongate cavities and skin-core structure in *Nephila* spider silk observed by electron microscopy. *J. Microsc.* **189**, 64–70 (1998).
 42. Work, R. W. Duality in major ampullate silk and precursive material from orb-web-building spiders (Araneae). *Trans. Am. Microsc. Soc.* **103**, 113–121 (1984).
 43. Vollrath, F., Holtet, T., Thøgersen, H. & Frische, S. Structural organization of spider silk. *Proc. R. Soc. Lond. B* **263**, 147–151 (1996).
 44. Viney, C., Huber, A. E., Dunaway, D. L., Kerkam, K. & Case, S. T. in *Silk Polymers. Materials Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 120–136 (American Chemical Society, Washington, 1994).
 45. Mahoney, D. V., Vezie, D. L., Eby, R. K., Adams, W. W. & Kaplan, D. in *Silk Polymers. Materials Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 196–210 (American Chemical Society, Washington, 1994).
 46. Shao, Z., Wen Hu, X., Frische, S. & Vollrath, F. Heterogeneous morphology in *Nephila edulis* spider silk and its significance for mechanical properties. *Polymers* **40**, 4709–4711 (1999).
 47. Hijirida, D. H., Do, K. G., Michal, C., Wong, S., Zax, D. & Jelinski, L. W. C13 NMR of *Nephila clavipes* major ampullate silk gland. *Biophys. J.* **71**, 3442–3447 (1996).
 48. Akai, H. The structure and ultrastructure of the silk gland. *Experientia* **39**, 443–449 (1983).
 49. Knight, D. P. & Vollrath, F. Liquid crystals and flow elongation in a spider's silk production line. *Proc. R. Soc. Lond. B* **266**, 519–523 (1999).
 50. Weiskopf, A., Senecal, K., Vouros, P., Kaplan, D. & Mello, C. M. The carbohydrate composition of spider silk: *Nephila edulis* dragline. *Glycobiology* **6**, 1703 (1996).
 51. Vollrath, F. & Tillinghast, E. K. Glycoprotein glue beneath a spiders web's aqueous coat. *Naturwissenschaften* **78**, 557–559 (1991).
 52. Kerkam, K., Viney, C., Kaplan, D. & Lombardi, S. Liquid crystallinity of natural silk secretions. *Nature* **349**, 596–598 (1991).
 53. Wilcox, P. J., Gido, S. P., Muller, W. & Kaplan, D. L. Evidence of a cholesteric liquid crystalline phase in natural silk spinning processes. *Macromolecules* **29**, 5106–5110 (1996).
 54. Magoshi, J., Magoshi, Y. & Nakamura, S. Crystallization, liquid crystal, and fiber formation of silk fibroin. *J. Appl. Polym. Sci.* **41**, 187–204 (1985).
 55. Nakamae, K., Nishino, T. & Ohkubo, H. Elastic modulus of the crystalline regions of silk fibroin. *Polymer* **30**, 1243–1246 (1989).
 56. Bunning, J. D. & Lydon, J. E. The cellular optical texture of the lyotropic nematic phase of the caesium pentadecafluoro-octanoate (ScPFO) water system in cylindrical tubes. *Liquid Cryst.* **20**, 381–385 (1996).
 57. Donald, A. M. & Windle, A. H. *Liquid Crystalline Polymers* 1–310 (Cambridge Univ. Press, Cambridge, 1992).
 58. Northolt, M. G. & Sikkema, D. J. Lyotropic main chain liquid crystal polymers. *Adv. Polym. Sci.* **98**, 115–177 (1991).
 59. Knight, D. P., Knight, M. M. & Vollrath, F. Beta transition and stress-induced phase separation in the spinning of spider dragline silk. *Int. J. Biol. Macromol.* **27**, 205–210 (2000).
 60. Riekel, C., Madsen, B., Knight, D. & Vollrath, F. Single fibre X-ray diffraction of spider silk. *Biol. Macromol.* (in the press).
 61. Palmer, J., Coyle, F. & Harrison, F. Structure and cyto-chemistry of the silk glands of the mygalomorph spider *Antrodiaetus unicolor* (Aranea, Antrodiaetidae). *J. Morphol.* **174**, 269–274 (1982).
 62. Tillinghast, E., Chase, S. & Townley, M. Water extraction by the major ampullate duct during silk formation in the spider, *Argiope aurantia* Lucas. *J. Insect Physiol.* **30**, 591–596 (1984).
 63. Vollrath, F., Madsen, B. & Shao, Z. The effect of spinning conditions on the mechanical properties of a spider's dragline. *Proc. R. Soc. Lond.* (in the press).
 64. Shultz, J. W. The origin of the spinning apparatus in spiders. *Biol. Rev.* **62**, 89–113 (1987).
 65. Tillinghast, E. K., Townley, M. A., Bernstein, D. T. & Gallagher, K. S. Comparative study of orb web hygroscopicity and adhesive spiral composition in three araneid spiders. *J. Exp. Zool.* **259**, 154–165 (1991).
 66. Schildknecht, H., Munzelmann, P., Krauss, D. & Kuhn, C. Über die chemie der spinnwebe. *Naturwissenschaften* **59**, 98–99 (1972).
 67. Vollrath, F. et al. Compounds in the droplets of the orb spider's viscid spiral. *Nature* **345**, 526–528 (1990).
 68. Edmonds, D. & Vollrath, F. The contribution of atmospheric water vapour to the formation and efficiency of a spider's web. *Proc. R. Soc. Lond.* **248**, 145–148 (1992).
 69. Vollrath, F. & Edmonds, D. Modulation of the mechanical properties of spider silk coating with water. *Nature* **340**, 305–307 (1989).
 70. Peters, H. M. Über den Spinnapparat von *Nephila madagascariensis* (Radnetzspinnen Argiopeidae). *Z. Naturforsch.* **10**, 395–404 (1955).
 71. Knight, D. P. & Vollrath, F. Changes in element composition along the spinning duct in a *Nephila* spider. *Naturwissenschaften* (in the press).
 72. Gosline, J. M., DeMont, M. E. & Denny, M. W. The structure and properties of spider silk. *Endeavour* **10**, 31–43 (1986).
 73. Vincent, J. F. V. *Structural Materials* 1–378 (Macmillan, London, 1982).
 74. Shao, Z. & Vollrath, F. The effect of solvents on the contraction and mechanical properties of spider silk. *Polymer* **40**, 1799–1806 (1999).
 75. Magoshi, J., Magoshi, Y. & Nakamura, S. Physical properties and structure of silk: 10. The mechanism of fibre formation from liquid silk of silkworm *Bombyx mori*. *Polym. Comm.* **26**, 309–311 (1985).
 76. Gamo, T. Genetic variants of the *Bombyx mori* silkworm encoding sericin proteins of different lengths. *Biochem. Genet.* **20**, 165–177 (1982).
 77. Sehnal, F. & Akai, H. Insect silk glands—their types, development and function, and effects of environmental-factors and morphogenetic hormones on them. *Int. J. Insect Morphol.* **19**, 79–132 (1990).
 78. Hepburn, H. R., Kurstjens, S. P. The combs of honeybees as composite-materials. *Apidologie* **19**, 25–36 (1988).
 79. Henderson, G., Manweiler, S. A., Lawrence, W. J., Tempelman, R. J. & Foil, L. D. The effects of *Steinernema capocapsae* (Weiser) application to different life stages on adult emergence of the cat flea *Ctenocephalides felis* (Bouche). *Vet. Dermatol.* **6**, 159–163 (1995).
 80. Zemlin, J. C. A *Study Of The Mechanical Behavior Of Spider Silks* 1–38 (Clothing and Organic Materials Laboratory Report, US Army Natick Labs, Natick, 1968).
 81. Engster, M. S. Studies on silk secretion in the Trichoptera (F. Limnephilidae). *Cell Tissue Res.* **169**, 77–92 (1976).
 82. Sivinski, J. Prey attraction by luminous larvae of the fungus gnat *Orfelia fultoni*. *Ecol. Entomol.* **7**, 443–446 (1982).
 83. Hunt, S. Amino acid composition of silk from the pseudoscorpion *Neobisium maritimum* (Leach): a possible link between the silk fibroins and the keratins. *Comp. Biochem. Physiol.* **34**, 773–777 (1970).
 84. Winkler, S. & Kaplan, D. L. Molecular biology of spider silk. *Rev. Mol. Biotech.* (in the press).
 85. Mello, C. M., Senecal, K., Yeung, B., Vouros, P. & Kaplan, D. in *Silk Polymers. Materials Science and Biotechnology* (eds Kaplan, D., Wade, W. W., Farmer, B. & Viney, C.) 67–79 (American Chemical Society, Washington, 1994).
 86. Candelas, G., Candelas, T., Ortiz, A. & Rodriguez, O. Translational pauses during a spider fibroin synthesis. *Biochem. Biophys. Res. Commun.* **116**, 1033–1038 (1983).
 87. van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935–938 (1999).
 88. van Beek, J. D., Kümmerlen, D., Vollrath, F. & Meier, B. H. Supercontracted spider dragline silk: a solid-state NMR study of the local structure. *Int. J. Biol. Macromol.* **24**, 173–178 (1999).
 89. Yeh, W.-Y. & Young, R. J. Molecular deformation processes in aromatic high modulus polymer fibres. *Polymer* **40**, 857–870 (1999).
 90. Shao, Z., Young, R. J. & Vollrath, F. The effects of solvents on spider silk studied by mechanical testing and single-fibre Raman spectroscopy. *Int. J. Biol. Macromol.* **24**, 295–300 (1999).
 91. Sirichaisit, S., Young, R. J. & Vollrath, F. Molecular deformation in spider dragline silk subjected to stress. *Polymers* **41**, 1223–1227 (1999).
 92. Riekel, C. et al. Aspects of x-ray diffraction on single spider fibers. *Macromolecule* **24**, 179–186 (1999).
 93. Grubb, D. T. & Jelinski, J. W. Fiber morphology of spider silk: the effects of tensile deformation. *Macromolecule* **30**, 2860–2867 (1997).
 94. Fung, Y. C. *Biomechanics* 1–586 (Springer, Heidelberg, 1981).
 95. Askeland, D. R. *The Science and Engineering of Materials* (PWS, Boston, 1994).
 96. Yoon, H. N., Charbonneau, L. F. & Calundann, G. W. Synthesis, processing and properties of thermotropic liquid-crystal polymers. *Adv. Materials* **4**, 206–214 (1992).
 97. Thiel, B. & Viney, C. A nonperiodic lattice model for crystals in *Nephila clavipes* major ampullate silk. *Mater. Res. Bull.* **20**, 52–56 (1995).

Acknowledgements

We thank B. Meier, R. Young, S. Case, J. Kenney, D. LaFollette, C. Craig, D. Kaplan, J. Gosline and H. Coulsey for their perceptive and helpful comments. Our research on spider silk has been funded by the Danish Science Research Council (SNF), the Science Faculty of Aarhus University, the Carlsberg Foundation, the Danish Academic Exchange Programme, the US Army (RDSG), the British Biological and Engineering Research Councils (BBSRC, EPSRC), the European Synchrotron Radiation Facility (ESRF) and the European Science Foundation (ESF). F.V. thanks the director and staff of Mpala Research Centre for their hospitality; D.P.K. thanks the Biological Imaging Centre at Southampton University for technical assistance and we both are grateful to our many collaborators for their help and support.

Correspondence and requests for materials should be addressed to F.V. (e-mail: fritz.vollrath@zoo.ox.ac.uk).

Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death

Nicholas Joza^{*,†}, Santos A. Susin[‡], Eric Daugas[§], William L. Stanford^{||}, Sarah K. Cho[†], Carol Y. J. Li^{||}, Takehiko Sasaki^{*§}, Andrew J. Elia^{*}, H.-Y. Mary Cheng^{*§}, Luigi Ravagnan[‡], Karine F. Ferri[‡], Naoufal Zamzami[‡], Andrew Wakeham^{*}, Razqallah Hakem^{*}, Hiroki Yoshida^{*}, Young-Yun Kong^{*}, Tak W. Mak^{*}, Juan Carlos Zúñiga-Pflücker[†], Guido Kroemer[‡] & Josef M. Penninger^{*†§}

* Amgen Institute, 620 University Avenue, Toronto, Ontario, Canada M5G 2C1

‡ Centre National de la Recherche Scientifique, UMR1599, Institut Gustave Roussy, 39 rue Calmette Desmoulins, F-94805 Villejuif, France

§ Assistance Publique, Hôpitaux de Paris, Service de Néphrologie B, Hôpital Tenon, 20 rue de la Chine, F-75020 Paris, France

|| Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5

† Ontario Cancer Institute, and the Departments of Medical Biophysics, and † Immunology, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Programmed cell death is a fundamental requirement for embryogenesis, organ metamorphosis and tissue homeostasis. In mammals, release of mitochondrial cytochrome *c* leads to the cytosolic assembly of the apoptosome—a caspase activation complex involving Apaf1 and caspase-9 that induces hallmarks of apoptosis. There are, however, mitochondrially regulated cell death pathways that are independent of Apaf1/caspase-9. We have previously cloned a molecule associated with programmed cell death called apoptosis-inducing factor (AIF). Like cytochrome *c*, AIF is localized to mitochondria and released in response to death stimuli. Here we show that genetic inactivation of AIF renders embryonic stem cells resistant to cell death after serum deprivation. Moreover, AIF is essential for programmed cell death during cavitation of embryoid bodies—the very first wave of cell death indispensable for mouse morphogenesis. AIF-dependent cell death displays structural features of apoptosis, and can be genetically uncoupled from Apaf1 and caspase-9 expression. Our data provide genetic evidence for a caspase-independent pathway of programmed cell death that controls early morphogenesis.

Programmed cell death (PCD) is a fundamental property of all multicellular organisms. It is crucial for plant and animal development, insect and amphibian metamorphosis, organ morphogenesis, tissue homeostasis, ageing, and the removal of infected or damaged cells¹. The biochemical and ultrastructural features of apoptosis are highly conserved throughout the evolution of multicellular animals^{1–4}. PCD has been linked to the CED9/Bcl-2, CED4/Apaf1 and CED3/caspase-9 genes that are essential for PCD in *Caenorhabditis elegans* and vertebrates^{5–8}. In response to death stimuli, mitochondrial membranes are permeabilized^{9,10}, and cytochrome *c* is released from mitochondria^{3,11,12} and associates with Apaf1 and pro-caspase-9 to trigger a caspase activation cascade that culminates in cell death characterized by apoptotic morphology^{7,13–15}. Failure to invoke appropriate cell death can result in cancer or autoimmunity, whereas increased PCD can lead to degenerative processes such as immunodeficiency and neurodegenerative disease¹⁶.

Although the cytochrome *c*/Apaf1/caspase-9 apoptosome is essential for several PCD pathways, cells deficient in these molecules can still die³. Indeed, cytochrome *c*, *apaf1* and *caspase-9* knockout mouse embryos undergo normal, albeit delayed, morphogenesis^{17–21}. Moreover, cell lines derived from these mutant mice are not uniformly resistant to death stimuli, but instead undergo PCD in a manner specific to both cell type and death signal¹⁹. It has also been shown that Bcl-2 preserves the integrity of mitochondrial membranes and protects cells from death independently of Apaf1 and caspases, implying that Bcl-2 interferes with two different mitochondrion-dependent death effector cascades^{22,23}. Thus, a death effector system other than cytochrome *c*/Apaf1/caspase-9 must be able to induce PCD.

We previously cloned apoptosis-inducing factor (AIF), which, like cytochrome *c*, is normally present in the mitochondrial intermembrane space and is released in response to death stimuli^{24,25}.

Extramitochondrial targeting of AIF, micro-injection of recombinant AIF protein into cells, or addition of AIF to isolated nuclei leads to the generation of apoptotic phenotypes, such as chromatin condensation and phosphatidylserine exposure on the cell surface²⁴. AIF has also been implicated in the control of apoptosis in syncytia induced by the HIV type-1 envelope glycoprotein²⁶, indicating that AIF may be involved in the pathogenesis of HIV infections. But although AIF can induce certain aspects of cell death in cultured cells, whether it is essential for PCD *in vivo* remains unresolved. Moreover, AIF has never been linked to PCD at the genetic level.

To explore the role of AIF in the control of PCD during animal development, we disrupted the mouse *aif* gene by homologous recombination. We report here that AIF is essential for the first wave of PCD required for embryonic morphogenesis and cavitation. Moreover, inactivation of AIF renders embryonic stem cells resistant to cell death after serum starvation. These results provide the first genetic evidence of a second, mitochondrially regulated cell death pathway in mammalian cells that is critical for morphogenesis and PCD after withdrawal of survival factors.

Gene targeting of *aif* in embryonic stem cells

The murine *aif* gene was ablated in embryonic stem (ES) cells using a targeting vector that deleted exon 3, corresponding to the amino terminus of the mature protein (nucleotides 247–346, amino acids 83–115). Three independent *aif*-targeted ES cell clones were obtained. Because the *aif* gene maps to the X chromosome²⁴, mutation of one *aif* allele resulted in a complete knockout in XY male ES cells and absence of *aif* expression by northern and western blotting (see Supplementary Information Fig. 1). As a control for changes to ES cells during G418 selection, ES cell clones were isolated in which the neomycin resistance cassette had integrated randomly into the genome (*aif*^{neo/Y}).

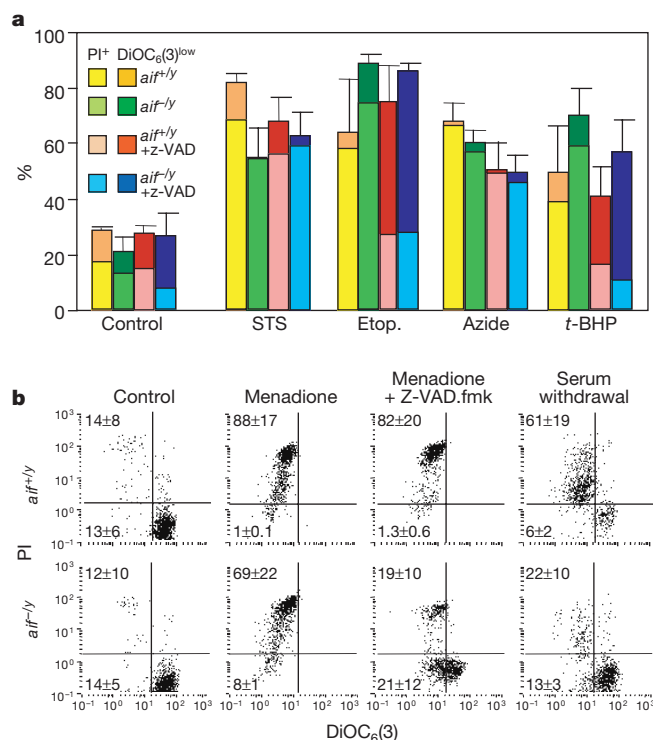


Figure 1 AIF is essential for cell death induced by serum withdrawal. **a**, Susceptibility of $aif^{-/y}$ ES cells to the death stimuli staurosporine (STS), etoposide (Etop.), azide and *tert*-butylhydroperoxide (t-BHP). ES cells were cultured in 10% serum in the presence of the indicated lethal stimuli and stained with propidium iodide (PI; for cell viability) and DiOC₆ (3) (for mitochondrial $\Delta\psi_m$). Note that all PI⁺ cells are DiOC₆ (3)^{low}. Mean values $\pm$ s.e.m. for three control and three $aif^{-/y}$ cell lines are shown. **b**, Resistance of $aif^{-/y}$ ES cells to cell death induced by menadione and serum withdrawal. FACS data are shown for $aif^{+/y}$ and a representative $aif^{-/y}$ ES cell clone. Numbers are mean values $\pm$ s.e.m. for three control and three different $aif^{-/y}$ cell lines in the corresponding quadrants.

Three independent $aif^{-/y}$ ES cell clones were injected into C57BL/6 blastocysts to generate chimaeric mice, and into $rag1^{-/-}$ blastocysts for lymphocyte reconstitution²⁷. Whereas all parental wild-type ES cell clones ($aif^{+/y}$) and all $aif^{neo/y}$ ES cell clones could contribute to adult tissues in chimaeric mice and reconstitute T- and B-cell lineages in $rag1^{-/-}$ mice, we failed to observe any chimaerism using all three $aif^{-/y}$ ES cells clones. Using *in vitro* ES cell differentiation and formation of teratocarcinoma-like tumours *in vivo*²⁸, however, $aif^{-/y}$ ES cell clones differentiated into cells from all three germ layers, including cartilage, muscle, neuronal tissue, epithelium, B cells, myeloid and erythroid cells (see Supplementary Information Fig. 1)²⁹. Thus, $aif^{-/y}$ ES cells retain their capacity to differentiate into cells from all three germ layers.

$aif^{-/y}$ ES cells are resistant to growth factor deprivation

The $aif^{-/y}$ ES cell lines exhibited normal proliferation *in vitro*. Unlike cytochrome $c^{-/-}$, $apaf1^{-/-}$ and $caspase-9^{-/-}$ ES cells^{17,19,21}, $aif^{-/y}$ ES cell lines displayed normal susceptibility to death, which was preceded by the dissipation of the mitochondrial transmembrane potential ($\Delta\psi_m$), in response to staurosporine, etoposide, azide, *tert*-butylhydroperoxide (Fig. 1a), anisomycin or ultraviolet irradiation (data not shown). This normal susceptibility to cell-death induction was observed both in the absence and in the presence of the pan-caspase inhibitor Z-VAD.fmk (Fig. 1a). Whereas serum withdrawal results in cell death of $aif^{+/y}$, $aif^{neo/y}$ and $apaf1^{-/-}$ ES cells²³, all three $aif^{-/y}$ ES cell lines largely conserved their viability and normal mitochondrial membrane integrity ($\Delta\psi_m$) when cultured in the absence of serum (Fig. 1b). Moreover, in the presence (but not the absence) of Z-VAD.fmk, $aif^{-/y}$ ES cell lines failed to die in response to the pro-apoptotic agent vitamin K₃ (menadione) (Fig. 1b). Thus, AIF is rate-limiting for some pathways of death induction. In particular, $aif^{-/y}$ ES cells are resistant to death after growth factor withdrawal.

AIF is essential for cavitation of embryoid bodies

The absence of overt chimaerism in whole organisms, yet the apparently normal differentiation potential of $aif^{-/y}$ ES cells *in vitro* and *in vivo*, suggested that AIF might be required for normal

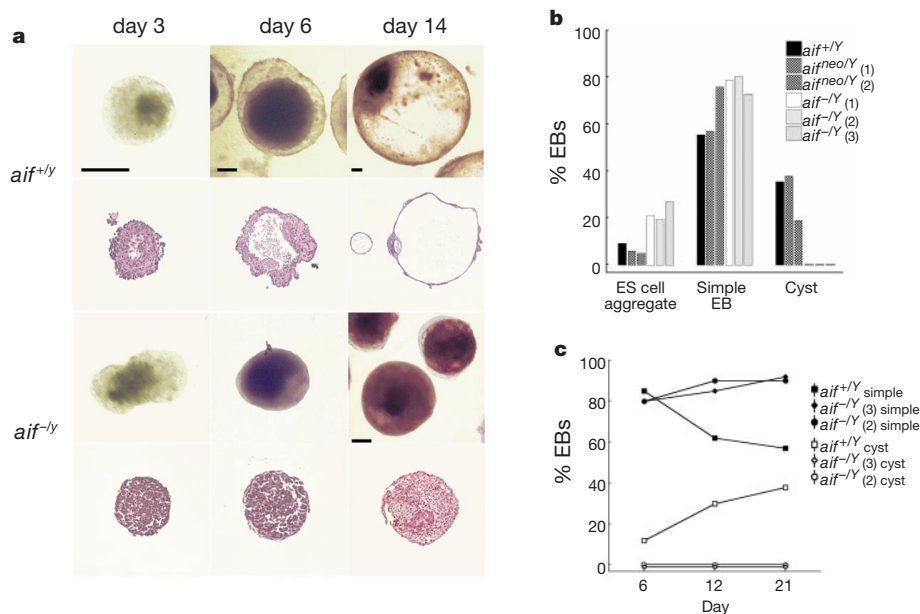


Figure 2 AIF is essential for embryoid body cavitation. **a**, Morphology (rows 1, 3) and histology (rows 2, 4; haematoxylin and eosin (HE) stain) of simple EBs (day 3), cystic (cavitated) EBs (day 6), and expanded cysts (day 14) derived from a wild-type ($aif^{+/y}$) and an $aif^{-/y}$ ES clone. Note the complete block in cavitation in the absence of AIF expression.

Scale bars, 100 μ m. **b**, Percentages at day 21 of ES cell aggregates, simple and cystic EBs from one $aif^{+/y}$, two $aif^{neo/y}$ and three $aif^{-/y}$ ES cell clones. At least 250 EBs were counted per genotype. **c**, Kinetics of EB formation. Shown is the frequency of simple and cystic EBs on days 6, 12 and 21.

PCD during early embryonic development. PCD occurs throughout mammalian development, beginning with apoptosis of the initially solid embryonic ectoderm to generate the proamniotic cavity³⁰. This early developmental process can be mimicked *in vitro* by culturing aggregates of ES cells in the absence of leukaemia inhibitory factor and feeder cells³¹. Under these culture conditions, ES cells form undifferentiated cell aggregates that develop into simple embryoid bodies (EBs), defined as multicellular aggregates containing an outer layer of endodermal cells and a solid core of undifferentiated ectodermal cells (Fig. 2a, left). The inner cells of simple EBs subsequently undergo PCD to form cystic EBs (Fig. 2a, top centre), a process called cavitation. As cystic embryoid bodies are cultured *in vitro*, the cavity expands (Fig. 2a, top right). The removal of cells of the inner core to form a cavitated or cystic EB is the first known wave of PCD during mouse morphogenesis³⁰.

When *aif*^{-/-} ES cells were tested in the EB formation assay, they were able to form simple EBs at frequencies and with kinetics comparable to those of *aif*^{+/+} and *aif*^{neo/+} controls (Fig. 2). But whereas a significant proportion of *aif*^{+/+} and *aif*^{neo/+} EBs underwent cavitation to form cystic EBs, EBs from all three differentiated *aif*^{-/-} ES cell lines exhibited a complete block in cavitation (Fig. 2a, b; and Supplementary Information Fig. 2). As cavitation is essential for the initiation of gastrulation and thus subsequent steps in embryogenesis³², defective cavitation by *aif*^{-/-} EBs probably explains the inability of *aif*^{-/-} ES cells to lead readily to adult tissue in chimaeric mice.

AIF controls PCD during early morphogenesis

Impaired cavitation might be due to either increased proliferation and/or impaired PCD of the cells that form the inner core. To

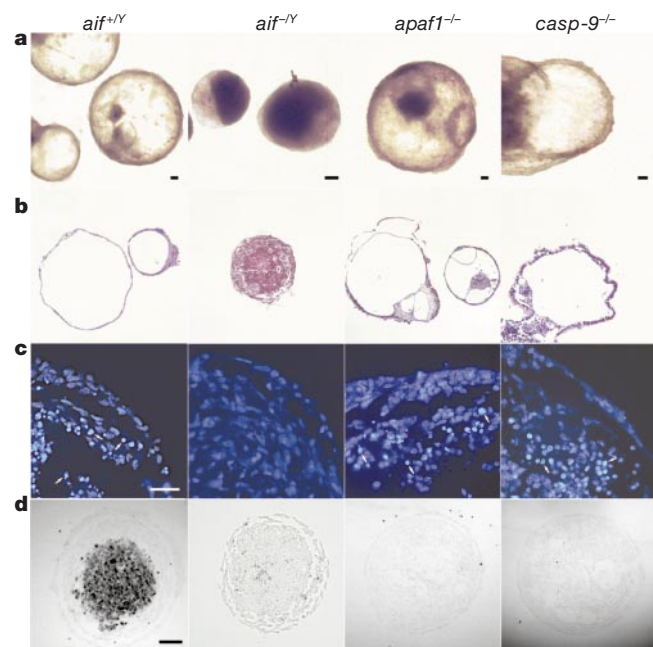


Figure 3 AIF is essential for PCD during early morphogenesis. **a, b**, Morphology (**a**) and histology (**b**; HE stain) of day 14 EBs from *aif*^{+/+} and *aif*^{-/-} ES cells (left), and *apaf1*^{-/-} and *caspase-9*^{-/-} ES cells (right). Note lack of cavitation in *aif*^{-/-} EBs but normal cavitation in *apaf1*^{-/-} and *caspase-9*^{-/-} EBs. Scale bars, 100 μm. **c, d**, Apoptotic features of inner cells from day 6 EBs from *aif*^{+/+} and *aif*^{-/-} ES cells (left), and *apaf1*^{-/-} and *caspase-9*^{-/-} ES cells (right). **c**, DAPI staining (blue) to visualize chromatin condensation (arrows). **d**, Caspase-3 activation. Note absence of chromatin condensation and caspase-3 activation in *aif*^{-/-} EBs. Similar results were obtained for two other *aif*^{-/-} ES clones and at later time points (up to day 21). All results for *aif*^{neo/+} EBs (*n* = 3) paralleled those obtained for wild-type EBs. Scale bar, 20 μm (**c**); 100 μm (**d**).

investigate whether EB cell proliferation was increased in the absence of AIF, we examined BrdU (5-bromodeoxyuridine) incorporation by wild-type and *aif*^{-/-} EBs. Although *aif*^{-/-} EBs displayed abnormal morphology (Fig. 3a, left) and histology (Fig. 3b, left), no evidence was obtained for increased proliferation of the inner cells of *aif*^{-/-} EBs at day 3, day 5, or at any later time point as compared with wild-type EBs (data not shown). To assay for PCD, the inner cells from wild-type and *aif*^{-/-} EBs were analysed by DAPI (4',6-diamidino-2-phenylindole dihydrochloride) staining to detect chromatin condensation (Fig. 3c, left) and by assays for *in situ* caspase-3 activation (Fig. 3d, left). Massive apoptosis was observed in the wild-type EBs, but no signs of cell death were found among the inner cells of *aif*^{-/-} EBs. These results indicate that impaired cavitation in *aif*^{-/-} EBs is not caused by enhanced proliferation but is due to a failure of inner cells to undergo PCD.

The outer endoderm cells have been suggested to provide death signals to inner cells required for cavitation³⁰; however, histological and electron microscopy analyses showed that simple *aif*^{-/-} EBs do not lack endodermal tissue. Furthermore, *aif*^{-/-} EBs expressed the endoderm-specific^{32,33} markers BMP2, BMP4, GATA-4, α -fetoprotein and HNF-4 (data not shown). To establish that the defects of *aif*^{-/-} cells are autonomous to inner cells, we generated chimaeric EBs by mixing *aif*^{-/-} and wild-type ES cells expressing a *lacZ* reporter (*aif*^{+/+}; *lacZ*) (Fig. 4a). Although cavitation was partially rescued in these chimaeric EBs (Fig. 4b), cell death was restricted to wild-type

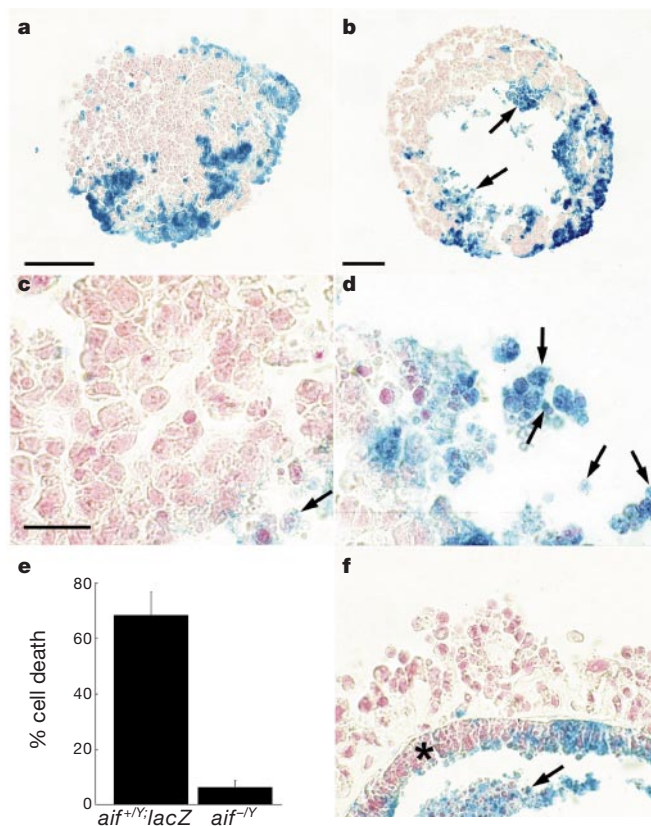


Figure 4 The effect of AIF is autonomous to inner cells. *aif*^{-/-} ES cells and wild-type cells expressing *lacZ* (*aif*^{+/+}; *lacZ*) were mixed to generate chimaeric EBs. **a–e**, At day 4 (**a**) or day 9 (**b–e**), EBs were stained with X-gal to mark *lacZ*-expressing cells (blue) and counterstained with nuclear fast red. **c, d**, High-magnification views of **b** showing viable *aif*^{-/-} inner cells (pink only) (**c**) and dead wild-type inner cells (blue) (**d**). **e**, Quantification of death in inner cells from day 9 chimaeric EBs. From each of 6 distinct EBs, 80–100 inner cells were counted. Mean values $\pm$ s.e.m. are shown. **f**, *aif*^{-/-} cells can differentiate into columnar epithelium (asterisk). Scale bar, 100 μm (**a, b**); 10 μm (**c**). Arrows in **b–d** and **f** indicate dead cells (fragmented nuclei and/or presence of cellular debris).

(blue) inner cells (Fig. 4b–e). Our mixing experiments also showed that *aif*^{-/-} cells can differentiate into columnar epithelium (Fig. 4f). These results indicate that impaired cavitation in *aif*^{-/-} EBs is not caused by defective endoderm formation. Instead, impaired cavitation is due to an intrinsic failure of AIF-deficient inner cells to undergo PCD.

Apaf1 and caspase-9 are not required for cavitation

The death of inner cells in wild-type EBs was found to be accompanied by the activation of caspase-3 (Fig. 3d, left), an effector caspase downstream of the cytochrome *c*/Apaf1/caspase-9 apoptosome. We therefore explored the contribution of Apaf1 and caspase-9 to cavitation by analysing the development of EBs from *apaf1*^{-/-} and *caspase-9*^{-/-} ES cells^{17,19}. Genetic inactivation of the *apaf1* and *caspase-9* genes abolished caspase-3 activation (Fig. 3d, right). However, loss of Apaf1 or caspase-9 expression had no apparent effect on cavitation (Fig. 3a, b; and Supplementary Information Fig. 3) or the death of inner cells (Fig. 3c). The kinetics and extent of cells that undergo chromatin condensation were comparable among wild-type, *apaf1*^{-/-} and *caspase-9*^{-/-} EBs (*n* = 5 per group). Adding the broad-spectrum caspase inhibitor z-VAD.fmk to developing wild-type EBs also failed to block cell death and subsequent cavitation. These results show that the PCD required for EB cavitation can occur in the absence of caspase-3 activation and can be genetically uncoupled from Apaf1 and caspase-9.

We next examined the intracellular localization of AIF in EBs and the effect of mutations of *apaf1*, *caspase-9* or *aif* on AIF and cytochrome *c* mobilization. In response to death stimuli AIF translocates from the mitochondria to the nucleus, whereas cytochrome *c* localizes to the cytosol²⁴. AIF (Fig. 5a, red colour) was found to translocate from the mitochondria to the nucleus (green DNA stain) in inner cells, but not in the outer endodermal cells of wild-type, *apaf1*^{-/-} and *caspase-9*^{-/-} EBs. Cytochrome *c* was also released from mitochondria of wild-type, *apaf1*^{-/-} and *caspase-9*^{-/-} inner cells (Fig. 5b). There was no detectable cytochrome *c* translocation from mitochondria to the cytosol in inner cells of *aif*^{-/-} EBs, indicating that mitochondrial membranes fail to permeabilize. This result is consistent with the failure of *aif*^{-/-} inner cells to activate caspase-3 (Fig. 3d), a defect that presumably results from deficient assembly of the apoptosome. These findings indicate that AIF acts upstream of cytochrome *c* and independently of the cytochrome *c*/Apaf1-triggered caspase activation cascade during cavitation.

AIF-regulated PCD has characteristic features of apoptosis

It has been reported that cell death of *apaf1*^{-/-} and *caspase-9*^{-/-} ES cells in response to ultraviolet radiation exhibits the morphological features of necrosis rather than apoptosis^{19,23}. To establish whether AIF-controlled cell death in EBs has the ultrastructural characteristics of apoptosis, inner cells from wild-type and *aif*^{-/-} EBs were compared using electron microscopy. Dying inner cells in wild-type

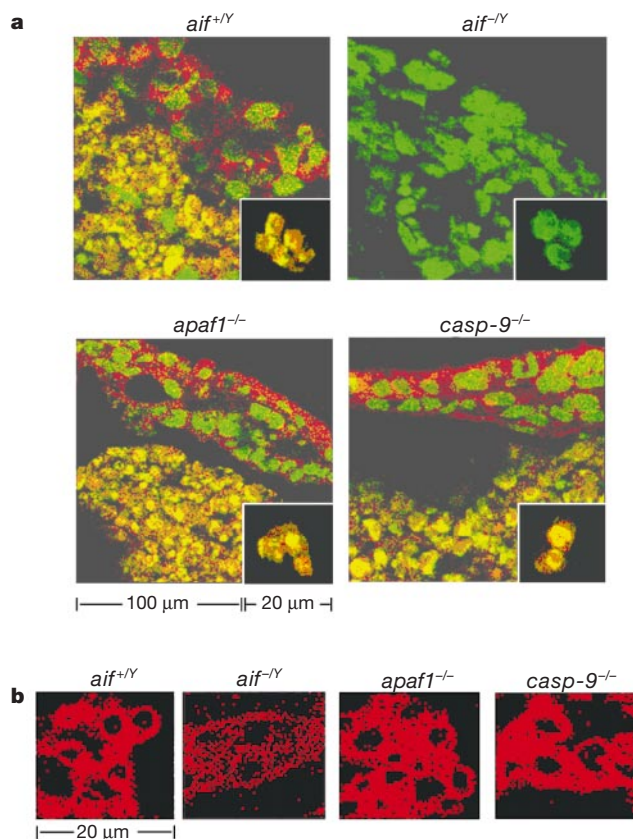


Figure 5 Translocation of AIF from the mitochondria to the nucleus. **a**, Immunolocalization of AIF (red) in outer and inner cells of day 6 wild-type, *apaf1*^{-/-}, *caspase-9*^{-/-} and *aif*^{-/-} EBs. In viable cells, AIF is sequestered in mitochondria (punctate red spots) separated from the nucleus (DNA-binding dye Sytox Green). In dying inner cells, AIF translocates from mitochondria to nuclei and green/red overlap appears yellow. Insets show high magnifications. **b**, Immunolocalization of cytochrome *c* (red) in inner cells of day 6 EBs. Cytochrome *c* (punctate staining in viable cells) accumulates in the cytosol in wild-type, *caspase-9*^{-/-} and *apaf1*^{-/-} cells undergoing PCD (diffuse staining). Cytochrome *c* is retained in mitochondria of *aif*^{-/-} cells.

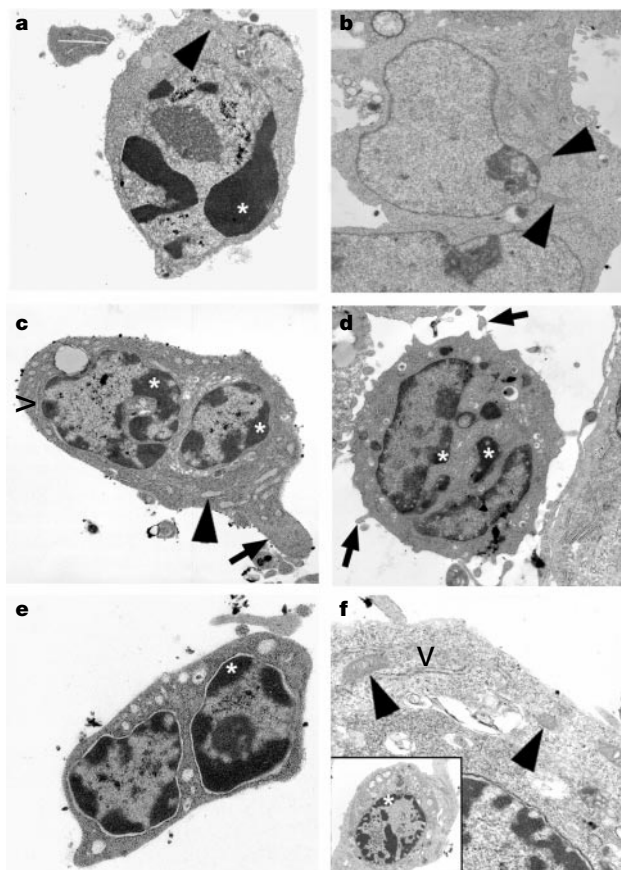


Figure 6 Morphological features of AIF-regulated PCD in EBs. Electron micrographs of inner ectodermal cells of day 6 wild-type (**a**), *aif*^{-/-} (**b**), *apaf1*^{-/-} (**c**) and *caspase-9*^{-/-} (**d–f**) EBs. A viable inner cell appears in *aif*^{-/-} EB (**b**). Note chromatin condensation (asterisks in **a** and **c–f**), plasma membrane blebbing (arrows in **c**, **d**), and preserved structural integrity of mitochondria (arrowheads in **a**, **c**, **f**) and of cytosolic organelles such as endoplasmic reticulum ('>' in **c**, **f**). **e**, Formation of apoptotic bodies. Original magnification, $\times 6,000$ (**a–d**); $\times 15,000$ (**e**); $\times 20,000$ (**f**).

EBs displayed typical apoptotic morphology (Fig. 6a), including the presence of chromatin condensation, plasma membrane blebbing, formation of apoptotic bodies, and a preserved ultrastructure of cytoplasmic organelles. Inner cells from *aif*^{-/-} EBs retained a healthy phenotype (Fig. 6b). Intriguingly, the inner cells from both *caspase-9*^{-/-} and *apaf1*^{-/-} EBs exhibited typical features of apoptosis, such as intact nuclear and plasma membranes, chromatin condensation (Fig. 6c–f, asterisks), plasma membrane blebbing (Fig. 6c, d, arrows), formation of apoptotic bodies (Fig. 6e), and preserved ultrastructure of mitochondria (Fig. 6c, f, solid arrowheads) and rough endoplasmic reticula (Fig. 6c, f, '>'). These features were similar to those in wild-type inner cells.

Consistent with the absence of caspase-3 activation, *caspase-9*^{-/-} and *apaf1*^{-/-} EBs do not manifest an advanced pattern of chromatin compaction (Fig. 6a). Instead, a peripheral type of chromatin compaction predominated (Fig. 6c–f). Thus, with the exception of caspase-dependent advanced chromatin compaction, the AIF-regulated pathway of PCD required for embryonic cavitation exhibits classical ultrastructural features of apoptosis and is independent of the Apaf1/caspase-9-mediated PCD pathway.

Discussion

In *C. elegans*, genetic evidence suggested that apoptosis is strictly dependent on caspase activation⁵. We provide genetic evidence here that not all apoptosis of mammalian cells is dependent on caspases, and that an AIF-dependent, caspase-independent PCD pathway exists that is crucial for cell death following growth factor deprivation and early mammalian development.

AIF and mitochondrial control of apoptosis. Numerous reports have shown that caspase inhibition prevents mammalian cell death or blocks the acquisition of morphological and biochemical characteristics of apoptosis³⁴. Moreover, mutational analyses of cytochrome *c* (ref. 21), caspases^{19,20} and Apaf1 (ref. 17) showed that these molecules contribute to apoptosis in a manner specific to both cell type and death signal. Similarly, *aif*^{-/-} ES cells, unlike *apaf1*^{-/-} and *caspase-9*^{-/-} ES cells, are sensitive to various apoptotic stimuli, such as staurosporine, anisomycin, ultraviolet irradiation and etoposide. However, *aif*^{-/-} ES cells are resistant to serum withdrawal and AIF is essential for the first wave of cell death during mouse morphogenesis. These data indicate the coexistence of two separate pathways linking the mitochondria to apoptosis, one that requires AIF and the other that relies on caspase activation.

The results of our study provide definitive genetic evidence that AIF inactivation abolishes all signs of cell death in early morphogenesis, including the mitochondrial release of cytochrome *c*. Moreover, AIF is a rate-limiting factor of ES cell death induced by menadione (only if caspases are simultaneously blocked) or by serum withdrawal (independently of caspase inhibition), indicating a stimulus-dependent contribution of AIF to the apoptotic cascade. The exact hierarchies and communication between AIF and the cytochrome *c*/Apaf1/caspase-9 apoptosome in cell-type- and death-signal-specific PCD remain to be determined.

Morphogenesis of multicellular organisms. PCD is essential during early animal development for the sculpting of digits, the palate and the eyes, the formation of hollow organs and the neural tube, and the generation of sexual organs¹. In early mouse embryos, the proamniotic cavity is formed by the death of the ectodermal cells in the core of the developing embryo³⁰. Thus, PCD is an integral part of morphogenesis and metamorphosis at all stages of animal development. Because developmental PCD exhibits the structural hallmarks of apoptosis, the finding that *C. elegans* bearing mutations of their caspase (CED-3) or Apaf1 (CED-4) orthologues have normal lifespans² was originally surprising. Moreover, morphogenesis and organ sculpting are also normal in cytochrome *c*, *apaf1* and *caspase-9* knockout mouse embryos, albeit delayed^{17–21}. These observations pointed to the existence of another PCD pathway that can compensate for the absence of caspase-dependent apoptosis and

that is highly conserved through evolution. As AIF messenger RNA and protein expression can be detected throughout murine embryogenesis and in all developing organs (see Supplementary Information Fig. 4), it is likely that AIF contributes to morphogenesis at later stages of embryogenesis.

We have shown that the genetic inactivation of AIF abolishes the first wave of developmental cell death occurring during early mouse embryogenesis. Assuming that ontogenesis recapitulates phylogeny, it is tempting to speculate that AIF represents a pathway of apoptosis that predates the caspase pathway. Whereas AIF homologues have been found in all metazoan phyla³⁵, no evidence for caspases has been reported in plants, fungi or unicellular organisms such as the *Trypanosoma cruzi* epimastigote, all of which can nevertheless undergo PCD¹. We propose that AIF and the AIF-regulated cell death pathway constitute an ancient and conserved process required for the morphogenesis of multicellular organisms. The identification of the molecules involved in this PCD pathway and their genetic and functional characterization should yield new insights into the basic physiology of cell death, and might allow us to develop strategies for the modulation of the cell death machinery. □

Methods

aif-deficient ES cells and chimaeric mice

The *aif* gene was cloned from a 129/SVJ mouse genomic library using a mouse *aif* probe (nucleotides 247–346). A targeting vector (600 base pairs short arm, and 6 kilobases long arm) flanking a PGK-Neo cassette was electroporated into male E14K ES cells. ES cell colonies resistant to G418 (300 µg ml⁻¹) were screened for homologous recombination by polymerase chain reaction (sense primer, 5'-GGGATTAGATAAATGCCTGCTCTT-3'; antisense primer, 5'-CCCCCAACTTATATCAGCCTACCTTC-3'). Recombinant colonies were confirmed by Southern blotting of *Hind*III-digested genomic DNA hybridized to a flanking probe. Total RNA was extracted from *aif*^{-/-} ES cells and subjected to northern blotting using full-length AIF complementary DNA. Absence of AIF protein in *aif*^{-/-} ES cells was determined by western blotting using an antibody reactive to residues 151–200 of murine AIF²⁴. Antibodies to Apaf1 (Upstate Biotechnology) and actin (Sigma) were used as controls. To test contribution to adult tissues, *aif*^{-/-} ES cells were injected into blastocysts from *rag1*^{-/-} mice²⁷ and C57BL/6 mice to generate chimaeric animals. Mice were maintained at the animal facilities of the Ontario Cancer Institute in accordance with institutional guidelines. Equivalent results and phenotypes were obtained for three independent *aif*^{-/-} ES cell clones. *Apaf1*^{-/-} and *caspase-9*^{-/-} ES cells have been described^{17,19}.

ES cell differentiation

Parental wild-type *aif*^{+/-}, three *aif*^{-/-} ES cell clones and ES cell clones in which Neo was randomly integrated (*aif*^{neo/y}) were cultured under conditions promoting differentiation into EBs^{29,36}. Chimaeric EBs were generated using *aif*^{-/-} ES cell clones (*lacZ*-negative) and *aif*^{+/-} ES cells constitutively expressing *lacZ* (*aif*^{+/-}; *lacZ*). The *aif*^{+/-}; *lacZ* ES cell clone contains a randomly integrated copy of the *lacZ* gene fused to the chicken β-actin promoter³⁷. For colony assays, single EB cell suspensions were replated in methylcellulose. Blood islands were detected using benzidine. ES cells were further differentiated into primitive mesodermal cells by co-culture with OP9 bone marrow stromal cells for 5 d. Single-cell suspensions from these cultures were either used for FACS analysis of Flk1⁺ hemangioblasts or replated onto OP9 cells and grown for an additional 5–12 d. Colonies were counted 10–14 d later and stained with Wright–Giemsa to analyse morphology, and with anti-CD45, CD11b, CD19 and TER-119 monoclonal antibodies³⁸. *In vivo* tumour formation of ES cells in athymic *nu/nu* mice and detection of differentiated tissues have been described²⁸.

Quantification of cell death

We cultured ES cells in the presence of 10% fetal calf serum and leukaemia inhibitory factor. Cell death was induced by addition of staurosporine (2 µM, 24 h), etoposide (100 µM, 24 h), sodium azide (15 µM, 48 h), *tert*-butylhydroperoxide (200 µM, 48 h) or menadione (150 µM, 24 h), or by serum withdrawal (0%, 72 h), in the presence or absence of Z-VAD.fmk (50 µM). Death was quantified by staining with propidium iodide (PI; 5 µg ml⁻¹) and DiOC₆(3) (40 nM).

Immunostaining and *in situ* procedures

For *in situ* localization of AIF and cytochrome *c* (ref. 25), paraformaldehyde-fixed EBs were stained with rabbit antiserum raised against residues 151–200 of AIF, or anti-cytochrome *c* monoclonal antibody (clone 6H2.B4, Pharmingen) followed by PE-conjugated goat anti-rabbit IgG (anti-AIF) or PE-conjugated goat anti-mouse IgG (anti-cytochrome *c*). Cells were counterstained with 10 nM Sytox Green (Molecular Probes). Staining was detected by confocal scanning fluorescence microscopy. Electron microscopy and *in situ* DAPI staining to detect chromatin condensation were as described¹⁹. Activated

caspase-3 was detected using an antibody specific for the cleaved (active) form of caspase-3 (New England BioLabs). *In situ* hybridization of murine embryos at distinct stages of development was done using sense and antisense probes from murine *aif* cDNA (nucleotides 456–1607).

Received 1 November 2000; accepted 31 January 2001.

- Jacobson, M. D., Weil, M. & Raff, M. C. Programmed cell death in animal development. *Cell* **88**, 347–354 (1997).
- Vaux, D. L. & Korsmeyer, S. J. Cell death in development. *Cell* **96**, 245–254 (1999).
- Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
- Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
- Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* **44**, 817–829 (1986).
- Hengartner, M. O. & Horvitz, H. R. *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* **76**, 665–676 (1994).
- Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome *c*-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
- Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. The *C. elegans* cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1 β -converting enzyme. *Cell* **75**, 641–652 (1993).
- Newmeyer, D. D., Farschon, D. M. & Reed, J. C. Cell-free apoptosis in *Xenopus* egg extracts: inhibition by Bcl-2 and requirement for an organelle fraction enriched in mitochondria. *Cell* **79**, 353–364 (1994).
- Kroemer, G. & Reed, J. C. Mitochondrial control of cell death. *Nature Med.* **6**, 513–519 (2000).
- Kluck, R. M., Bossy-Wetzel, E., Green, D. R. & Newmeyer, D. D. The release of cytochrome *c* from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science* **275**, 1132–1136 (1997).
- Yang, J. *et al.* Prevention of apoptosis by Bcl-2: release of cytochrome *c* from mitochondria blocked. *Science* **275**, 1129–1132 (1997).
- Liu, X., Kim, C. N., Yang, J., Jemmerson, R. & Wang, X. Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome *c*. *Cell* **86**, 147–157 (1996).
- Li, P. *et al.* Cytochrome *c* and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
- Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1/cytochrome *c* multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
- Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1462 (1995).
- Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
- Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
- Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways *in vivo*. *Cell* **94**, 339–352 (1998).
- Kuida, K. *et al.* Reduced apoptosis and cytochrome *c*-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
- Li, K. *et al.* Cytochrome *c* deficiency causes embryonic lethality and attenuates stress-induced apoptosis. *Cell* **101**, 389–399 (2000).
- Amarante-Mendes, G. P. *et al.* Anti-apoptotic oncogenes prevent caspase-dependent and independent commitment for cell death. *Cell Death Differ.* **5**, 298–306 (1998).
- Haraguchi, M. *et al.* Apoptotic protease activating factor 1 (Apaf-1)-independent cell death suppression by Bcl-2. *J. Exp. Med.* **191**, 1709–1720 (2000).

- Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**, 441–446 (1999).
- Daugas, E. *et al.* Mitochondrio-nuclear translocation of AIF in apoptosis and necrosis. *FASEB J.* **14**, 729–739 (2000).
- Ferri, K. F. *et al.* Apoptosis control in syncytia induced by the HIV type 1-envelope glycoprotein complex. Role of mitochondria and caspases. *J. Exp. Med.* **192**, 1081–1092 (2000).
- Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
- Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
- Bautch, V. L. *et al.* Blood island formation in attached cultures of murine embryonic stem cells. *Dev. Dyn.* **205**, 1–12 (1996).
- Coucouvanis, E. & Martin, G. R. Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* **83**, 279–287 (1995).
- Robertson, E. J. *Embryo-Derived Stem Cell Lines* (IRL, Oxford, 1987).
- Coucouvanis, E. & Martin, G. R. BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development* **126**, 535–546 (1999).
- Leahy, A., Xiong, J. W., Kuhnert, F. & Stuhlmann, H. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. *J. Exp. Zool.* **284**, 67–81 (1999).
- McCarthy, N. J., Whyte, M. K., Gilbert, C. S. & Evan, G. I. Inhibition of Ced-3/ICE-related proteases does not prevent cell death induced by oncogenes, DNA damage, or the Bcl-2 homologue Bak. *J. Cell Biol.* **136**, 215–227 (1997).
- Lorenzo, H. K., Susin, S. A., Penninger, J. & Kroemer, G. Apoptosis inducing factor (AIF): a phylogenetically old, caspase-independent effector of cell death. *Cell Death Differ.* **6**, 516–524 (1999).
- Schmitt, R. M., Bruyns, E. & Snodgrass, H. R. Hematopoietic development of embryonic stem cells *in vitro*: cytokine and receptor gene expression. *Genes Dev.* **5**, 728–740 (1991).
- Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
- Cho, S. K. *et al.* Functional characterization of B lymphocytes generated *in vitro* from embryonic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 9797–9802 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Saunders for scientific editing; A. Oliveira-dos-Santos, K. Bachmaier, T. Wada, V. Stambolic, L. Zhang, M. Crackower, C. Krawczyk, I. Kozieradzki, Q. Liu, J. Irie-Sasaki, M. Nghiem, R. Sarao, E. Griffith, L. Barra and A. Manoukian for comments; D. Métié and B. Calvieri for technical assistance; and J. Rossant and A. Bernstein for *lacZ*-expressing ES cells. N.J. and J.M.P. are supported by the Canadian Institute of Health Research (CIHR), Amgen, and the National Cancer Institute of Canada. W.L.S. is supported by the Karyn Glick Memorial Special Fellowship and CIHR. E.D. is supported by Assistance Publique-Hôpitaux de Paris and CANAM. G.K. is supported by grants from Ligue Nationale Contre le Cancer, European Commission and Agence Nationale pour la Recherche Sur le SIDA. J.M.P. holds a Canadian Research Chair in Cell Biology.

Correspondence and requests for materials should be addressed to J.M.P. (e-mail: jpennin@amgen.com).

Production of iron nanoparticles by laser irradiation in a simulation of lunar-like space weathering

Sho Sasaki*, Keiko Nakamura†, Yoshimi Hamabe*, Erika Kurahashi* & Takahiro Hiroi‡

* Department of Earth and Planetary Science, The University of Tokyo, Tokyo 113-0033, Japan

† Department of Earth and Planetary Sciences, Kobe University, Kobe 657-8501, Japan

‡ Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

'Space weathering' is the term applied to the darkening and reddening of planetary surface materials with time, along with the changes to the depths of absorption bands in their optical spectra. It has been invoked to explain the mismatched spectra of lunar rocks and regolith, and between those of asteroids and meteorites^{1–6}. The formation of nanophase iron particles on regolith grains as a result of micrometeorite impacts or irradiation by the solar wind has been proposed as the main cause of the change in the optical properties^{7,8}. But laboratory simulations^{9–14} have not revealed the presence of these particles, although nano-second-pulse laser irradiation did reproduce the optical changes¹². Here we report observations by transmission electron microscopy of olivine samples subjected to pulse laser irradiation. We find within the amorphous vapour-deposited rims of olivine grains nanophase iron particles similar to those observed in the rims of space-weathered lunar regolith grains^{15,16}. Reduction by hydrogen atoms implanted by the solar wind is therefore not necessary to form the particles. Moreover, the results support the idea that ordinary chondrites came from S-type asteroids⁵, and thereby provides some constraints on the surface exposure ages of those asteroids.

Space weathering was originally proposed to explain the spectral difference between lunar soils and underlying rocks, and it might also explain spectral mismatches between asteroid types and meteorite classes, especially S-type asteroids and ordinary chondrites. Lunar soil studies indicated that space weathering may be caused by the accumulation of the dark agglutinitic glass formed by melting of regolith by meteorite impacts¹. However, the optical changes may principally be caused⁷ by nanophase iron particles (about 10 nm in size) within coatings produced by recondensation of ferrous silicate vapour. Impact heating of micrometeorites and interplanetary dust particles is a process that could plausibly produce the vapour, although the sputtering and implantation of solar wind particles could also cause the surface alteration. Recently, a theoretical scattering model has shown that nanophase iron particles on grain surface rims should darken and redden the reflectance spectra of ordinary chondrites and lunar soils⁸. In fact, nanophase iron particles were found on the vapour-deposited rims of lunar soil grains^{15,16}. An important consortium study on lunar soils⁶ has shown that space weathering on the Moon cannot be explained solely by agglutinitic glasses: the formation of nanophase iron particles may also control the change of optical properties, especially reddening.

Previously, heating experiments have been performed to simulate space weathering. Spectral changes after the melting of meteorite samples by a fusion furnace were measured but the changes were caused by glass formation⁹. Using a pulse laser as a heating device, changes of reflectance spectra were observed¹⁰. However, the laser pulse duration was 0.5–1 μ s, which was 1,000 times longer than the feasible timescale of micrometeorite (1–10 μ m size) impacts. The

laser-treated samples were molten and the resultant spectral changes should be ascribed to glass formation.

In order to simulate space weathering by high-velocity dust impact, we irradiated pellet samples of olivine and pyroxene with a pulse laser beam (1,064 nm) under a vacuum at $(1–2) \times 10^{-5}$ torr. The pulse duration was 6–8 ns, which is comparable to the timescale of dust impacts^{11,12}. Laser-irradiated samples showed significant reddening: reduction of reflectance was much larger in the visible wavelength region than in the near-infrared. Reflectances of olivines are more easily changed¹¹ than those of pyroxenes; this result is compatible with compiled reflectance data on asteroids^{13,14}. Some of the visible–near-infrared asteroid spectra can be reproduced¹⁴ by mixing spectra of laser-irradiated olivine and pyroxene samples.

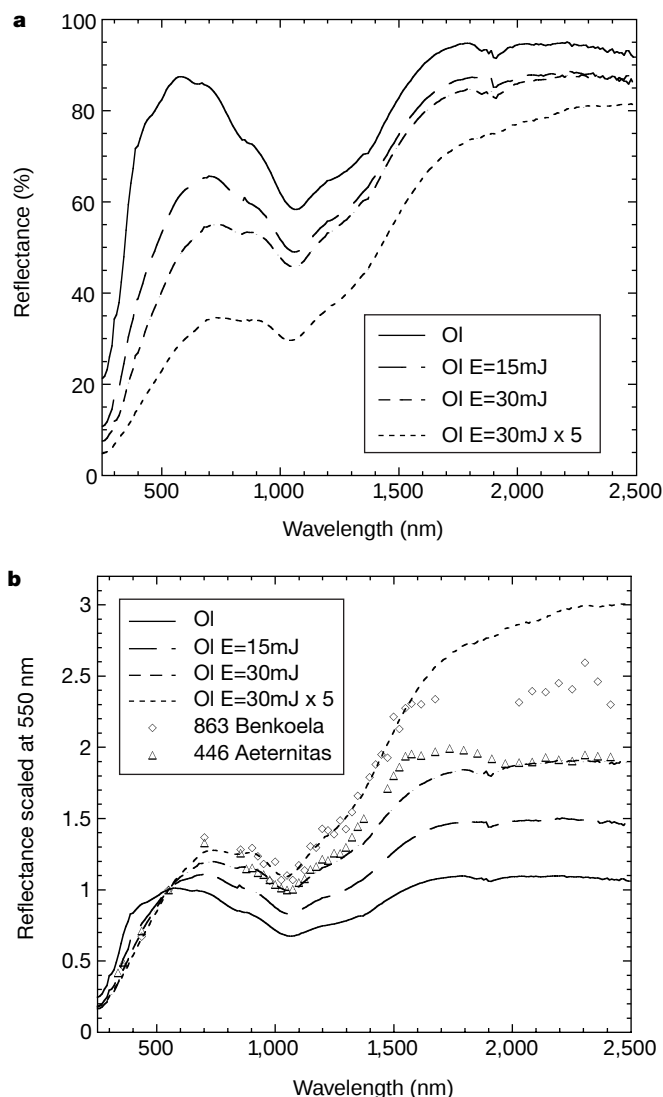


Figure 1 Bidirectional reflectance spectra of olivine pellet samples before and after pulse laser irradiation. Each 2-cm pellet is made up of grains smaller than 75 μ m. The area uniformly irradiated by pulse laser has dimensions 1 cm $\times$ 1 cm, and the central millimetre-sized region of the pellet is carefully selected for measurements of bidirectional reflectance spectrum. Spectra are measured at the facility installed at the Tsukuba Space Center of the National Space Development Agency of Japan; a spectrum range between 250 and 2,500 nm is recorded at every 10 nm. **a**, Absolute spectra of olivine. **b**, Olivine spectra scaled at 550 nm. There are four spectra (non-irradiated, 15-mJ irradiation, 30-mJ irradiation, and five repetitions of 30-mJ irradiation treatment). The scaled spectra are compared with A-type asteroids 446 Aeternitas and 863 Benkoela which are considered to belong to an olivine-rich clan of S-type because of their lack of the 2- μ m absorption band.

To visualize the microscopic effects causing reflectance changes, we performed repetitive pulse laser irradiations on olivine pellet samples and observed the irradiated samples using transmission electron microscopy (TEM) after obtaining reflectance measurements. We use San Carlos Olivine with 8.97 wt% FeO ($\text{Fe}_{0.1}$). Reflectance spectra of raw and laser-irradiated samples are shown in Fig. 1. In the absolute spectrum (Fig. 1a), reflectance at 500 nm decreased by 51% after 30 mJ irradiation and by as much as 67% after a five-repetition irradiation treatment of 30 mJ. Compared with the large depletion in visible reflectance, infrared reflectance decreased by only 20–30% even after five repetitions of 30-mJ irradiation. As a result, after repeated irradiations, the spectrum becomes more reddened, which is evident in the scaled spectra (Fig. 1b). The absolute depth of the 1- μm absorption band becomes shallower but the 1- μm band still remains detectable in the scaled spectrum. In Fig. 1b, spectra are compared with the observed spectra of asteroids which are considered to be olivine-rich¹⁷. In particular the 446 Aeternitas spectrum is close to the olivine spectra after 30-mJ irradiation. The scaled spectra of irradiated olivine are compatible with the flattening of S-asteroid spectra at wavelengths longward of 1.5 μm ^{18,19}. The more reddened slope of the 863 Benkoela spectrum is reproduced by that of a five-repetition 30-mJ irradiation treatment at wavelengths shorter than 1.5 μm , although repeated irradiation produced continuously red spectrum at longer wavelengths.

Several grains from each olivine pellet sample after being irradiated five times or twenty times were thin-sectioned for TEM

measurement. The initial olivine San Carlos sample was also prepared for the comparable investigation. The rim regions of the irradiated olivine samples (amorphous rims of $\sim 200\text{-nm}$ thickness) are developed along the grain rims. In high-resolution TEM images, nanophase particles (several up to 30 nm in size) are recognized that are widely dispersed throughout the amorphous rims (Fig. 2). These particles appear to be single- to poly-crystallines embedded within the amorphous material. Neither nanophase particles nor amorphous rims were observed on initial unirradiated olivine samples. The electron diffraction patterns from these crystallines show typical $\alpha\text{-Fe}$ patterns (body-centred cubic, the unit length $a_0 = 0.2874\text{ nm}$). In the high-resolution TEM images, we observe a regular 0.204-nm basal spacing which is consistent with the spacing of a crystal lattice plane (110) of $\alpha\text{-Fe}$ ($d_{110} = a_0/\sqrt{2} = 0.203\text{ nm}$). The energy-dispersive X-ray analyses on the regions where these nano-phase particles densely occur indicate that the beam-indent area is enriched in iron. Average atomic percentages of the amorphous materials of the rims from six different areas are: O, 59.05 (57.07); Si, 18.10 (14.08); Mg, 20.60 (25.05); and Fe, 2.25 (3.80); where the values in brackets are compositional data from the host olivine. The amorphous material is more silica-rich than the underlying olivine crystal. Moreover, as seen in Fig. 2b, doubled layers in rim structures are frequently observed. We may conclude that the amorphous silica rim with $\alpha\text{-Fe}$ nano-particles is produced by vapour deposition. Our observed $\alpha\text{-Fe}$ nanophase particles are clearly similar to those found in the rim of lunar soil grains in occurrence and size^{15,16}.

Our results suggest that we have succeeded in simulating a lunar-like space weathering process. We confirm a suspected link between the occurrence of nanophase iron in vapour-deposited grain rims and the spectroscopic optical effects of darkening and reddening. In space, heating by micrometeorite impacts on the regolith surface should produce the vapour, part of which would subsequently deposit on nearby grains. Through this process, nanophase crystalline iron particles should be formed and deposited. Our study shows

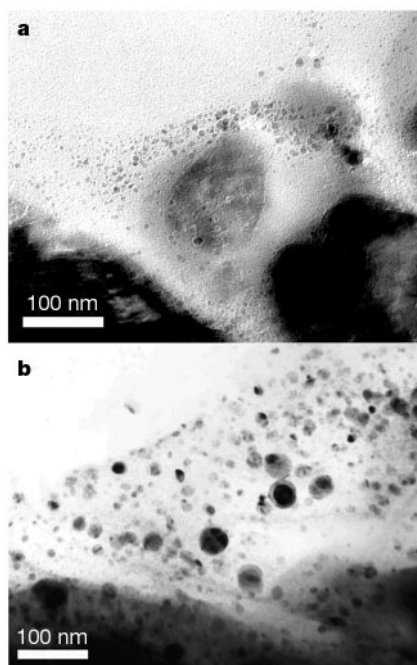


Figure 2 Nanophase iron particles in bright-field TEM images of the rim of laser-irradiated olivine grains. Selected grains from pellet samples were embedded within low viscosity epoxy (Quetol812) and thin-sectioned with $\sim 80\text{ nm}$ thickness using an ultramicrotome. Imaging and electron diffraction were performed with a JEOL2010 (200 keV) TEM equipped with an energy-dispersive X-ray spectrometer at Kobe University. The electron diffraction was calibrated by using a gold-coated holey carbon film. **a**, Amorphous rim of an olivine grain from a sample irradiated 5 times at 30 mJ. There are many particles with size ranging from several nanometres to 30 nm. The boundary between host olivine crystal and amorphous rim is sharp. **b**, Doubled, possibly multiple, amorphous layers of a sample irradiated 20 times at 30 mJ. The outer rim is more widely developed than both the inner rim and the rim of the five times irradiated samples. We note that nanophase particles in the outer rim are slightly larger than those in the inner rim and polycrystallines are more ubiquitous in the outer rim.

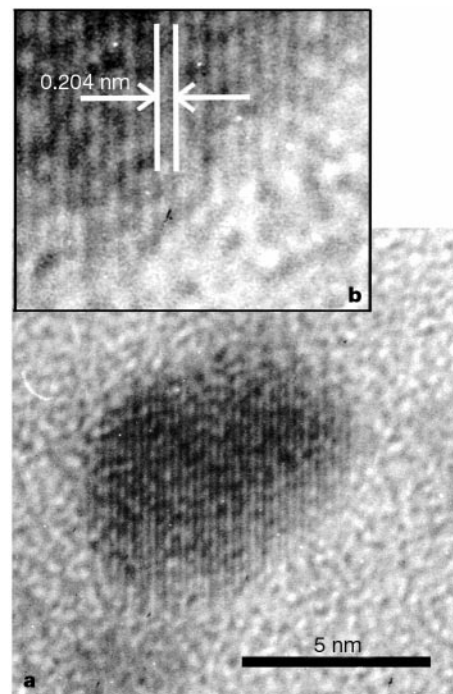


Figure 3 A nanophase particle in the rim of an olivine grain irradiated 20 times at 30 mJ. **a**, A bright-field high-resolution TEM image. The grain size is 7.7 nm. **b**, An enlarged image of **a**. The measured interlayer spacing is 0.203–0.205 nm, which is consistent with the spacing of a crystal lattice plane (110) of $\alpha\text{-Fe}$ ($d_{110} = 0.203\text{ nm}$).

that reduction by implanted hydrogen from solar winds is not necessary to form nanophase iron, although a previous study found formation of micrometre-scale or smaller iron particles only under hydrogen-induced reducing conditions with continuous heating²⁰.

Our results are compatible with recent discoveries, supporting the idea that space weathering has changed the optical properties of S-type asteroids. Although S-type asteroids are the most common in the inner part of the main asteroid belt, steep reddened reflectance spectra and derived mineralogies of S-type asteroids are different from those of ordinary chondrites, the most common meteorites^{2–4}. Many small near-Earth asteroids have intermediate spectra between S-type and Q-type (ordinary-chondrite like) spectra²¹. Multispectral observation of an S-type body Ida by the Galileo spacecraft demonstrated that relatively fresh crater interiors or ejecta show colour properties similar to ordinary chondrites⁵. Furthermore, recent measurement by an X-ray spectrometer on board the NEAR-Shoemaker spacecraft showed that S-type asteroid 433 Eros has an elemental composition close to unfractionated ordinary chondrites²². Here we have shown that a physical process (micrometeorite impact heating) should link both with production of nanophase iron particles (Figs 2 and 3) and with spectral changes (Fig. 1) similar to those that would change ordinary chondrite spectra into S-type spectra.

Because the physical process is simulated by pulse laser heating, we can tentatively constrain the exposure ages of asteroidal regolith surfaces. The energy deposition rate in our laser experiments is $2.4 \times 10^5 \text{ J m}^{-2}$ at 30-mJ pulse laser irradiation¹¹. In space around 1 AU, the impact rate of dust particles of around 10^{-12} g ($1 \mu\text{m}$ in diameter) is a few $10^{-4} \text{ m}^{-2} \text{ s}^{-1}$ (ref. 23). Assuming a relatively high impact velocity of around 20 km s^{-1} for vaporization, the energy release rate by dust impacts is about $10^{-3} \text{ J m}^{-2} \text{ yr}^{-1}$. When efficiencies used for heating/vaporization are the same for both laser-irradiated energy and dust-impact energy, irradiation of a 30-mJ pulse laser in our experiments corresponds to dust-impact heating over about 10^8 yr in space. Comparisons between asteroidal and laser-irradiated spectra may constrain the surface exposure age of asteroids. The energy efficiency at real micrometeorite impacts is unknown; if it is lower than that at laser irradiation as we expect, this age might be the lower limit. Our results can be applied to constrain the relative ages of optically fresh and optically mature regoliths on a single asteroid. However, care should be taken in making age comparisons between different asteroids, as the degree of space weathering should depend on the chemical composition, such as the olivine-to-pyroxene ratio, and also on the heliocentric distance, which affects the micrometeorite flux. □

Received 15 November 2000; accepted 26 January 2001.

- Adams, J. B. & McCord, T. B. Alteration of lunar optical properties: age and composition effects. *Science* **171**, 567–571 (1971).
- Chapman, C. R. & Salisbury, J. W. Comparison of meteorite and asteroid spectral reflectivities. *Icarus* **19**, 507–522 (1973).
- Gaffey, M. J. & McCord, T. B. Asteroid surface materials: Mineralogical characterizations from reflectance spectra. *Space Sci. Rev.* **21**, 555–628 (1978).
- Wetherill, G. W. & Chapman, C. R. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 35–67 (Univ. Arizona Press, Tucson, 1988).
- Chapman, C. R. S-type asteroids, ordinary chondrites, and space weathering: The evidence from Galileo's fly-bys of Gaspra and Ida. *Meteor. Planet. Sci.* **31**, 699–725 (1996).
- Pieters, C. M. *et al.* Space weathering on airless bodies: Resolving a mystery with lunar samples. *Meteor. Planet. Sci.* **35**, 1101–1107 (2000).
- Hapke, B., Cassidy, W. & Wells, E. Effects of vapor-phase deposition processes on the optical, chemical and magnetic properties of the lunar regolith. *Moon* **13**, 339–353 (1975).
- Hapke, B. Space weathering in the asteroid belt. *Lunar Planet. Sci.* **31**, 1087 (2000).
- Clark, B. E., Fanale, F. P. & Salisbury, J. W. Meteorite-asteroid spectral comparison: The effects of comminution, melting, and recrystallization. *Icarus* **97**, 288–297 (1992).
- Moroz, L. V., Fisenko, A. V., Semjonova, L. F., Pieters, C. M. & Korotaeva, N. N. Optical effects of regolith processes on S-asteroids as simulated by laser shots on ordinary chondrite and other mafic materials. *Icarus* **122**, 366–382 (1996).
- Yamada, M. *et al.* in *Antarctic Meteorites Vol. 23*, 173–176 (National Inst. Polar Res., Tokyo, 1998).
- Yamada, M. *et al.* Simulation of space weathering of planet-forming materials: Nanosecond pulse laser irradiation and proton implantation on olivine and pyroxene samples. *Earth Planets Space* **51**, 1255–1265 (1999).

- Hiroi, T. & Sasaki, S. Importance of olivine in S-asteroid space weathering. *Lunar Planet. Sci.* **30**, 1444 (1999).
- Hiroi, T. & Sasaki, S. First successful simulation of the asteroid space weathering using reflectance spectra of pulse laser irradiated olivine and pyroxene samples—Possible compositional dependency of S-asteroid space weathering. *Meteor. Planet. Sci.* (submitted).
- Keller, L. P. & McKay, D. S. Discovery of vapor deposits in the lunar regolith. *Science* **261**, 1305–1307 (1993).
- Keller, L. P. & McKay, D. S. The nature and origin of rims on lunar soil grains. *Geochim. Cosmochim. Acta* **61**, 2331–2341 (1997).
- Gaffey, M. J. *et al.* Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993).
- Clark, B. E. & Hiroi, T. S-type asteroid spectral continua: Redness and Fe, Ni metal. *Bull. Am. Astron. Soc.* **26**, 1172 (1994).
- Clark, B. E. Spectral mixing models of S-type asteroids. *J. Geophys. Res.* **100**, 14443–14456 (1995).
- Allen, C. C., Morris, R. V., Lauer, H. V. Jr. & McKay, D. S. Microscopic iron metal on glass and minerals—A tool for studying regolith maturity. *Icarus* **104**, 291–300 (1993).
- Binzel, R. P., Bus, S. J., Burbine, T. H. & Sunshine, J. M. Spectral properties of near-Earth asteroids: Evidence for sources of ordinary chondrite meteorites. *Science* **273**, 946–948 (1996).
- Trombka, J. I. *et al.* The elemental composition of asteroid 433 Eros: Results of the NEAR-Shoemaker X-ray spectrometer. *Science* **289**, 2101–2105 (2000).
- Grün, E. *et al.* in *Origin and Evolution of Interplanetary Dust* (eds Levasseur-Regourd, A. C. & Hasegawa, H.) 21–32 (Kluwer, Dordrecht, 1991).

Acknowledgements

We thank M. Yamada for discussions and help in laser irradiation experiments, J. Owada and H. Akiyama for technical support during reflectance measurements, and H. Nagahara and K. Tomeoka for discussions and support. We thank T. Kogure for preliminary measurements by TEM and K. Akamatsu for technical advice during our TEM measurements. We also thank L. Keller for discussions on TEM analysis. We thank C. R. Chapman and B. E. Clark for their valuable comments. This work was partially supported by a Grant-in-Aid for Science Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan and by a grant from the Inamori Foundation.

Correspondence and requests for materials should be addressed to S.S.
(e-mail: sho@eps.s.u-tokyo.ac.jp).

Rapid magnetic reconnection in the Earth's magnetosphere mediated by whistler waves

X. H. Deng & H. Matsumoto

Radio Science Center for Space and Atmosphere, Kyoto University, Uji, Kyoto 611-0011, Japan

Magnetic reconnection has a crucial role in a variety of plasma environments^{1–3} in providing a mechanism for the fast release of stored magnetic energy. During reconnection the plasma forms a 'magnetic nozzle', like the nozzle of a hose, and the rate is controlled by how fast plasma can flow out of the nozzle. But the traditional picture of reconnection has been unable to explain satisfactorily the short timescales associated with the energy release, because the flow is mediated by heavy ions with a slow resultant velocity. Recent theoretical work^{4–6} has suggested that the energy release is instead mediated by electrons in waves called 'whistlers', which move much faster for a given perturbation of the magnetic field because of their smaller mass. Moreover, the whistler velocity and associated plasma velocity both increase as the 'nozzle' becomes narrower. A narrower nozzle therefore no longer reduces the total plasma flow—the outflow is independent of the size of the nozzle. Here we report observations demonstrating that reconnection in the magnetosphere is driven by whistlers, in good agreement with the theoretical predictions.

The topological change occurring during magnetic reconnection requires that the ideal magnetohydrodynamics (MHD) constraints must be broken. But the magnetosphere is essentially collisionless, like the solar corona. Traditionally it has been believed that Alfvén

waves played a key role in driving reconnection: the waves accelerate the plasma out of the nozzle up to a maximum, which is the Alfvén velocity. The Alfvén waves are perturbations of the magnetic field lines, which are controlled by the motions of ions, but the reconnection rate in the conventional resistive MHD model is much too small to explain the timescale of energy release unless a large localized or current-dependent resistivity is assumed. However, there has been no evidence of such anomalous resistivity. Recently a new theory was proposed claiming that the dissipation region develops a distinct multi-scale structure in which the dynamics at small-scale lengths are controlled by whistler-mode waves^{4–6}. These ‘whistlers’ are responsible for accelerating the plasma out of the magnetic nozzle during reconnection. On the inertial scale, the motion of electrons and of ions decouples, and the magnetic field is frozen only to the electron fluid. Whistlers are basically supported by the motion of electrons, with much lighter mass, and hence have a much faster velocity for small-scale perturbations. The Hall current caused by the relative motion of electrons and ions produces a transition from Alfvén-wave-driven reconnection to whistler-mediated reconnection⁷. The whistlers therefore replace Alfvén waves when the nozzle becomes sufficiently narrow. The reconnection rate, in contrast, becomes large and insensitive to the specific mechanism that breaks the frozen-in condition, and remains almost the same even at very large scales^{5,6}. The internal generation of the out-of-plane magnetic field pattern B_y , caused by the Hall current—and illustrated in Fig. 1—is regarded as the main signature of whistler-mediated reconnection.

The well-known coronal mass ejection hit the Earth on 10 January 1997, making the current layer thickness become comparable to the ion skin length⁸. During this period the GEOTAIL spacecraft skimmed along the dayside magnetopause. Figure 2 shows a summary of the observation by GEOTAIL for the interval 05:35–06:05 UT.

The first piece of evidence of the Hall effect is the appearance of the out-of-plane B_y having typical localized bipolar structures associated with the passage of magnetic islands identified by the bipolar signatures in B_x . There was a distinct increase in B_y (Fig. 2c) occurring almost simultaneously with the B_z decrease (Fig. 2d), while GEOTAIL approached the low-latitude boundary layer at about 05:39 UT. By comparing the GEOTAIL data with the data of the Wind spacecraft in interplanetary space, we identified that the B_y increase comes from interplanetary space but was largely amplified by the bow shock effect and enhanced by the compression of the twisted magnetic fields. Note that in Fig. 2b there was a series of bipolar signatures in B_x , which is a basic feature of flux transfer events⁹. Figure 2a illustrates the B_x and B_y on an extended timescale from 05:40 to 05:45 UT, showing a series of bipolar variations in B_y around the time-averaged value of B_y . They are well correlated with the bipolar signatures of B_x (identified as I_1 to I_7). For example, at the I_2 event, GEOTAIL stayed south of the magnetic equator and on the magnetosheath side of the magnetopause. Thus, when a magnetic island passed through GEOTAIL, we observed a bipolar signature in B_x changing from negative (earthward) to positive (sunward). Synchronizing with the bipolar signature of B_x , the variation in B_y was recorded, first decreasing and then increasing around the time-averaged value of B_y . Although the out-of-plane B_y is driven by the electron-scale flows, it broadens to a region comparable to the ion inertial length¹⁰. Cross-correlation analysis shows that the coefficient of B_x and B_y is largest when there is no time lag between them. This supports the physical picture shown in Fig. 1. The series of bipolar signatures of B_y associated with the magnetic island structures can be well explained by the quadrupole nature of B_y caused by the electron Hall current^{11,12} as demonstrated in Fig. 1.

The second piece of evidence for a Hall current is the observation of the polarity change of the out-of-plane B_y during multiple traverses of the magnetopause boundary; these encounters are

due to the motion of the magnetopause and are identified by the polarity change of B_z (identified as P_1 to P_6 in Fig. 3). They are well correlated with the change of the polarity of B_y . Taking the P_2 and P_3 events as an example, during the period of P_2 GEOTAIL passed across the magnetopause (Fig. 3b) from the magnetosheath to the magnetosphere, and recorded the variation in B_y (Fig. 3a), first decreasing and then increasing around the time-averaged value of B_y . During the period of P_3 , GEOTAIL traversed the magnetopause from the magnetosphere to the magnetosheath, recording a variation in B_y , first increasing and then decreasing around the time-averaged value of B_y . The change in the polarity of B_y during the several magnetopause encounters can be well explained by the quadrupole structure of B_y caused by the electron Hall current as demonstrated in Fig. 1.

The third piece of evidence supporting whistler-mediated reconnection is the observation of the enhancement of electromagnetic waves with frequencies in the range of oblique whistler-mode waves. By surveying the data of GEOTAIL skimming along the dayside magnetopause, we find a good correlation between magnetic reconnection and the enhancement of wave activities. Figure 2f shows the dynamic frequency spectrogram of electric and magnetic

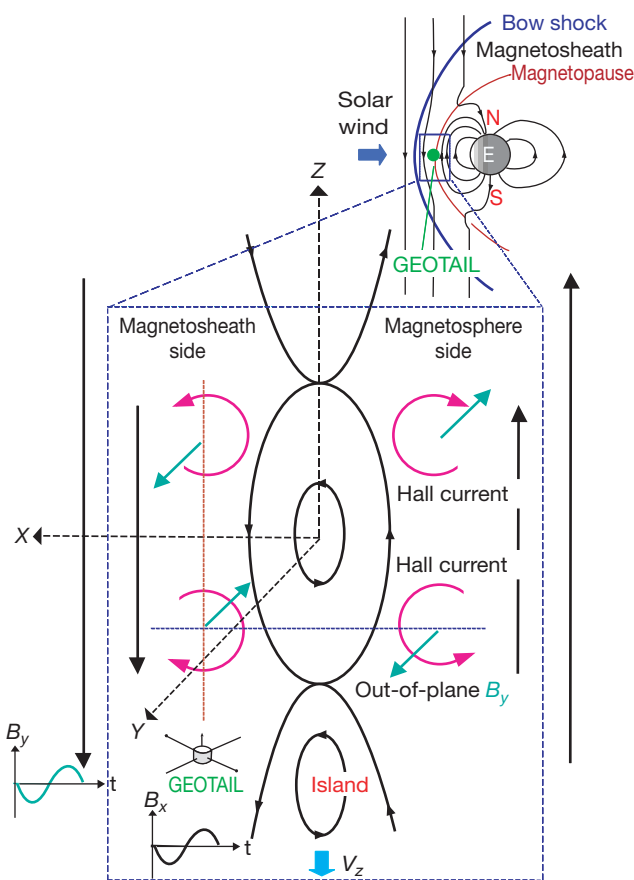


Figure 1 Collisionless reconnection configuration at dayside magnetopause for 10 January 1997. The GSE (geocentric solar equatorial) coordinate system is used, with x towards the Sun and z northwards. Solid black lines denote magnetic field lines. Green arrows present the self-generated quadrupole out-of-plane magnetic field component of B_y due to Hall currents, which are denoted by thick pink encircling lines. The electron inertial term and the Hall term have associated scale lengths of electron inertial length δ_e and ion inertial length δ_i , where $\delta_e = c/\omega_{pe}$ and $\delta_i = c/\omega_{pi}$, c being the speed of light, and ω_{pe} and ω_{pi} being the electron and ion plasma frequencies respectively. The dynamical response of electrons to distortions of the magnetic field in the unmagnetized ion limit is described by whistler-mode waves rather than Alfvén waves. Whistler-mode waves are transverse right-handed circularly polarized waves. Their electric vector rotates in the same sense as electron cyclotron motion around the external magnetic field.

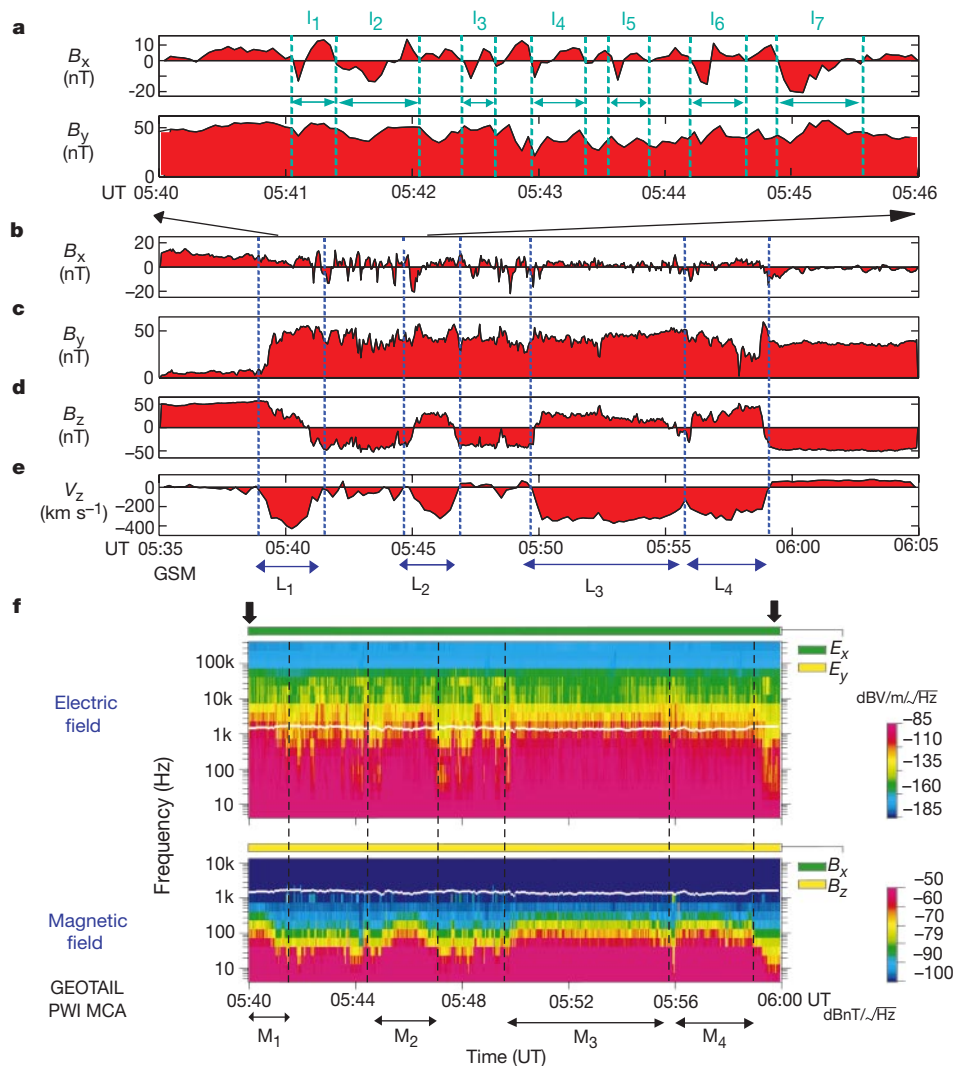


Figure 2 Plasma and magnetic field data and dynamic frequency spectrogram of electric and magnetic fields observed by GEOTAIL PWI-MCA (Multi-Channel Analyzer) on transverse through the dayside magnetopause boundary on 10 January 1997. **b–e**, The magnetic field components of B_x , B_y and B_z (**b–d**) and the plasma flow speed component of V_z . **a**, B_x and B_y on an extended timescale. **f**, The frequency–time spectrogram of electric and magnetic fields. The white line shows the local electron cyclotron frequency. Under the parameters near the magnetopause (plasma density $n = 10 \text{ cm}^{-3}$, total

magnetic field $B_T = 60 \text{ nT}$ and electron temperature $T_e = 100 \text{ eV}$), the ion inertial length ($\delta_i = 72 \text{ km}$) is comparable to the characteristic current layer thickness. From **a**, the averaged temporal scale of the bipolar signatures of B_x and B_y is $\sim 20 \text{ s}$. Whereas the magnetic islands are convected away from the reconnection region with an average speed of $30\text{--}60 \text{ km s}^{-1}$, the spatial scale of B_x and B_y is $\sim 600\text{--}1,200 \text{ km}$, which is comparable to the scale length of the order of $10\delta_i$ of the theoretical prediction.

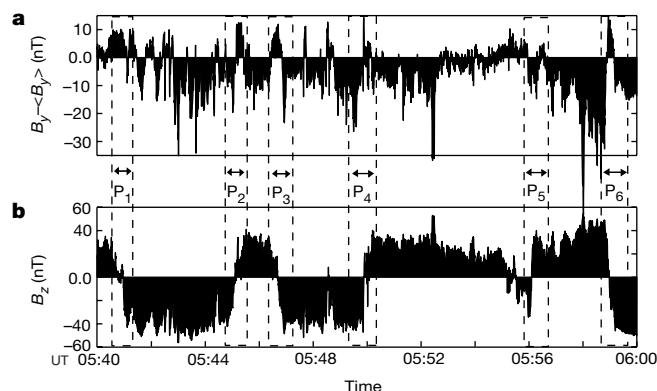


Figure 3 High-resolution magnetic field data for GEOTAIL magnetopause crossing on 10 January 1997. **a**, **b**, Magnetic field components of B_y (**a**) and B_z (**b**). The magnetopause encounters (P_1 to P_6) are identified by the changes in the field orientation of B_z . The crossings are also confirmed by changes in temperature and density (data not

shown). Initially GEOTAIL was located in the magnetosphere region (southwards B_z with a negative value) in the Southern Hemisphere. Owing to boundary motions, GEOTAIL experienced multiple magnetopause encounters before it finally left towards the magnetosheath side (northwards B_z with a positive value).

fields measured by the plasma wave instrument (PWI)^{13,15}. It shows that four bursts of enhanced wave activities (M_1 , M_2 , M_3 and M_4) were observed synchronously with the four velocity spikes of accelerated plasma flow V_z (L_1 , L_2 , L_3 and L_4 in Fig. 2e). Accelerated plasma flow is a typical signature of magnetic reconnection¹⁴. A burst of the electromagnetic waves with frequencies in the range between the electron cyclotron frequency and the lower hybrid frequency was recorded by waveform capture of the PWI¹⁵ at 05:39 UT (data not shown). The wavevector $\mathbf{k}$ turns out to be nearly perpendicular to the ambient magnetic field by minimum variance analysis. The polarity is right-handed in the plane perpendicular to the ambient magnetic field, which is typical of electron whistler-mode waves. Measurement of broadband electric and magnetic fluctuations peaking in the current layer, with amplitudes proportional to the local magnetic shear angle, provides observational evidence of enhanced whistler-mode fluctuations at the magnetopause and supports the current as the source of the fluctuations^{16,17}.

The above-mentioned observations support the theory of whistler-mediated reconnection by revealing the existence of the internally generated quadrupole structure of the out-of-plane B_y . The Hall effect is the critical factor in collisionless magnetic reconnection, and the dispersion property of whistler-mode waves is the key to understanding the result that the reconnection rate is independent of any mechanism that breaks the frozen-in flux condition^{4–6}. However, the electron current layers of scale length δ_e and the density depletion layer demonstrated in simulations^{10,18} have not been observed. It has been shown^{4,18,19} that the electron current layers are unstable to electron shear flow instability, lower-hybrid drift instability and kinetic kink instability. Perturbations developing as a result of these instabilities around the reconnection region break up the electron current layer and can wash out the density depletion layer. This is a major step in understanding the absence of such narrow layers²⁰. However, further work is needed to establish the nature of the turbulence, which instability has the major and crucial role in triggering reconnection, and how particles are accelerated. □

Received 7 June; accepted 12 December 2000.

- Paschmann, G. *et al.* Plasma acceleration at the earth's magnetopause: Evidence for reconnection. *Nature* **282**, 243–246 (1979).
- Priest, E. R. *Solar Flare Magnetohydrodynamics* (Gordon & Breach, London, 1981).
- Urrutia, J. M., Stenzel, R. L. & Rousculp, C. L. Pulsed currents carried by whistlers. III. Magnetic fields and currents excited by an electrode. *Phys. Plasmas* **2**, 1100–1113 (1995).
- Drake, J. F., Biskamp, D. & Zeiler, A. Breakup of the electron current layer during 3-D collisionless magnetic reconnection. *Geophys. Res. Lett.* **24**, 2921–2924 (1997).
- Birn, J. *et al.* GEM magnetic reconnection challenge. *J. Geophys. Res.* (in the press).
- Shay, M. A., Drake, J. F., Rogers, B. N. & Denton, R. E. The scaling of collisionless magnetic reconnection for large systems. *Geophys. Res. Lett.* **26**, 2163–2166 (1999).
- Mandt, M. E., Denton, R. E. & Drake, J. F. Transition to whistler mediated magnetic reconnection. *Geophys. Res. Lett.* **21**, 73–76 (1994).
- Le, G. & Russell, C. T. The thickness and structure of high beta magnetopause current layer. *Geophys. Res. Lett.* **21**, 2451–2454 (1994).
- Lee, L. C. & Fu, Z. F. A theory of magnetic flux transfer at the Earth's magnetopause. *Geophys. Res. Lett.* **12**, 105–108 (1985).
- Shay, M. A., Drake, J. F., Rogers, B. N. & Denton, R. E. Alfvénic collisionless magnetic reconnection and the Hall term. *J. Geophys. Res.* (in the press).
- Sonnerup, B. U. O. in *Solar System Plasma Physics* (eds Lanzerotti, L. J., Kennel, C. F. & Parker, E. N.) Vol. 3, 45–108 (North-Holland, Amsterdam, 1979).
- Drake, J. F., Kleva, R. G. & Mandt, M. E. Structure of thin current layers: Implications for magnetic reconnection. *Phys. Rev. Lett.* **73**, 1251–1254 (1994).
- Anderson, R. R. *et al.* Plasma waves near the magnetopause. *J. Geophys. Res.* **87**, 2087–2107 (1982).
- Petschek, H. E. in *AAS/NASA Symposium on the Physics of Solar Flares* (ed. Hess, W. N.) 425–439 (NASA Spec. Publ. SP-50, Washington DC, 1964).
- Matsumoto, H. *et al.* Plasma wave observations with GEOTAIL spacecraft. *J. Geomagn. Geoelectr.* **46**, 59–95 (1994).
- Zhu, Z. *et al.* The relationship between ELF-VHF waves and magnetic shear at the dayside magnetopause. *Geophys. Res. Lett.* **23**, 773–776 (1996).
- Song, P. *et al.* Properties of ELF emissions in the dayside magnetopause. *J. Geophys. Res.* **103**, 26495–26506 (1998).
- Hesse, M., Schindler, K., Birn, J. & Kuznetsova, M. The diffusion region in collisionless magnetic reconnection. *Phys. Plasmas* **5**, 1781–1795 (1999).
- Pritchett, P. L., Coroniti, F. V. & Decyk, V. K. Three-dimensional stability of thin quasi-neutral current sheets. *J. Geophys. Res.* **101**, 27413–27429 (1996).
- Russell, C. T. & Elphic, R. C. ISEE observations of flux transfer events at the dayside magnetopause. *Geophys. Res. Lett.* **6**, 33–36 (1979).

Acknowledgements

We thank H. Kojima, J. F. Drake and R. R. Anderson for their help and valuable discussion, and T. Mukai and N. Nagai for providing the particle and magnetic field data, respectively. D.X.H. is on leave from Wuhan University and received support from the RASC Visiting Foreign Scientist programme. This work was supported by a Grant-in-Aid for Scientific Research.

Correspondence and requests for materials should be addressed to H.M. (e-mail: matsumot@kurasc.kyoto-u.ac.jp).

Manipulation of elementary charge in a silicon charge-coupled device

Akira Fujiwara & Yasuo Takahashi

NTT Basic Research Laboratories, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan

The ultimate limit in the operation of an electronic device is the manipulation of a single charge. Such a limit has been achieved in single-electron tunnelling devices^{1,2}. However, these devices are based on multiple tunnel barriers and conductive islands, which are complex structures to fabricate. Here we demonstrate another type of device that can also manipulate elementary charge, but which is more suitable for large-scale integration. The device consists of two closely packed silicon wire-MOSFETs, which are commonly used building blocks of electronic circuits. We have developed a scheme to generate and store holes in the channels of either of these MOSFETs. Subsequently, holes can be transferred between the two MOSFETs at the level of an elementary charge, and their exact position can be monitored. This single-charge transfer device, which is operated at 25 K, is in effect a charge-coupled device³. This is also the first realization of a silicon-based device that manipulates elementary charge.

Single-electron devices^{1,2} have been attracting much attention, owing to their underlying physics and the application to future integrated circuits⁴. Their operation is based on the Coulomb blockade⁵; the tunnelling of an electron into an ultrasmall conductive island is inhibited by the charging energy. Metal was the first material to be used; now various materials such as the compound

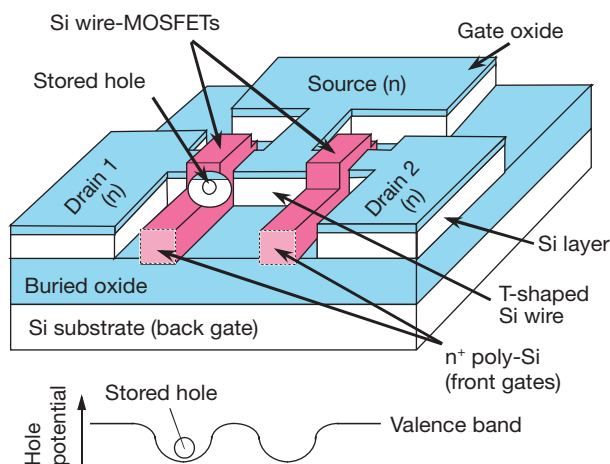


Figure 1 Schematic view of a single-electron CCD on a silicon-on-insulator wafer. The Si wire-MOSFETs are defined by the fine poly-Si gates that cover a part of the T-shaped Si wire. In addition to the fine gates, the upper poly-Si gate is formed (after the formation of SiO_2 interlayer) to cover a large area (not shown here). The Si wire is surrounded by gate oxide (not shown). A single hole can be stored and transferred.

semiconductors⁶, Si⁷ and carbon nanotubes⁸ have been researched, particularly a device consisting of one island. Such a device is the SET (single-electron tunnelling) transistor⁹ that works as a current switch. There have been many reports on SET transistors using metals^{10,11}, GaAs^{12,13}, and Si^{14–16}. Single-electron memories have also been widely investigated (metal¹⁷ and Si^{18–20}).

The manipulation of single electrons is possible in the single-electron transfer devices such as single-electron turnstiles²¹ and single-electron pumps²², which include coupled plural islands. In these devices, one electron can be precisely transferred from one electrode to another. They can be applied to current standards²³ in metrology and future circuits where one-bit information is related to the single electron²⁴. However, fabrication is difficult because it requires the integration of multiple islands and tunnel capacitors. Single-electron transfer has not yet been performed in Si devices^{25,26}.

A charge-coupled device (CCD)³—CCDs are now widely used in image sensors—consists of an array of MOS diodes. A suitable sequence of clock voltage pulses to the diodes traps bundles of electrons in a potential well and transfers them from one diode to another. We expect that the scaling-down of CCDs will finally lead to single-electron manipulation. The importance of CCDs as single-electron transfer devices has been pointed out⁴; they are suitable for high-frequency operation because the electron transfer mechanism is not tunnelling.

We fabricated an ultrasmall Si CCD. Figure 1 shows a schematic diagram of the device fabricated by electron beam lithography on a silicon-on-insulator^{27,28} wafer. It is a closely packed array of two Si wire-MOSFETs. The channels of the MOSFETs are defined by fine n+ poly-Si gates (90 nm long). Single holes can be stored under these gates (see below) and transferred back and forth between them. The T-shaped Si wire shared by the two MOSFETs is connected to three large n-type electrodes (one source and two drains). These electrodes are used to make electron currents flow. These currents are important in sensing stored holes, which will be described later.

The estimated thicknesses of the Si wire and the gate oxide are 15–20 nm and 35–40 nm, respectively. We used double-layer structures for the poly-Si gates. After the formation of an SiO₂ interlayer, the upper gate is formed (both not shown in Fig. 1). The upper gate covers a large area and is used as a mask for ion implantation to form n+ source and drain regions. The fine n+ poly-Si gates act as front

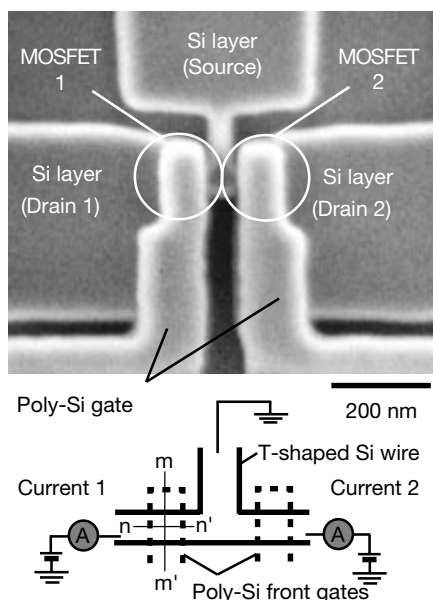


Figure 2 Top-view SEM image and simplified circuit diagram of the device. The wide upper gate (sense gate) is not shown. The electron currents flowing through the two wire-MOSFETs can be measured separately.

gates and the Si substrate acts as a backgate. We use the upper gate to make the electron channel open and name this gate the 'sense gate'. A top-view SEM image of the device and the circuit diagram is shown in Fig. 2. If we make the device operate as an n-channel MOSFET, the source is grounded and the drain voltages are positively biased; thus two electron currents can be detected separately.

We used an unusual method of charge storage and sensing. We wanted to separate electrons and holes and locate them at different Si/SiO₂ interfaces of the Si wire. This was achieved by applying a large electric field across the Si wire. Figure 3a shows schematic cross-sections of the Si wire, which are along the m–m' and n–n' line in Fig. 2. A large negative voltage is applied to the front gate and a large positive voltage is applied to the back gate. To store holes in the Si wire, we illuminate the device for a few seconds with a halogen lamp or a He–Ne laser. This 'one-shot illumination' generates holes that are localized at the front interface owing to the large electric field. Then, to sense the holes, the sense-gate voltage is scanned in the positive direction while the front and back gate voltages are kept constant. Then, at a certain voltage, electrons are induced at the back interface and the n-channel current flows. Because the electron channel potential is affected by the Coulomb potential by the stored holes, we can calculate the amount of stored holes from the current level. If we regard the wire-MOSFET as a hole-trapping island, its capacitance is estimated to be 11 aF. Then, the potential change by a single hole should be 15 meV. This is large enough for detecting the holes with the sensitivity of an elementary charge at 25 K.

Figure 3b shows the drain current versus the sense-gate voltage at 25 K for one wire-MOSFET. The front-gate voltage and the back-gate voltage were fixed at –1.6 V and 60 V, respectively. The drain voltage was 100 mV. We can crudely estimate the electric field across

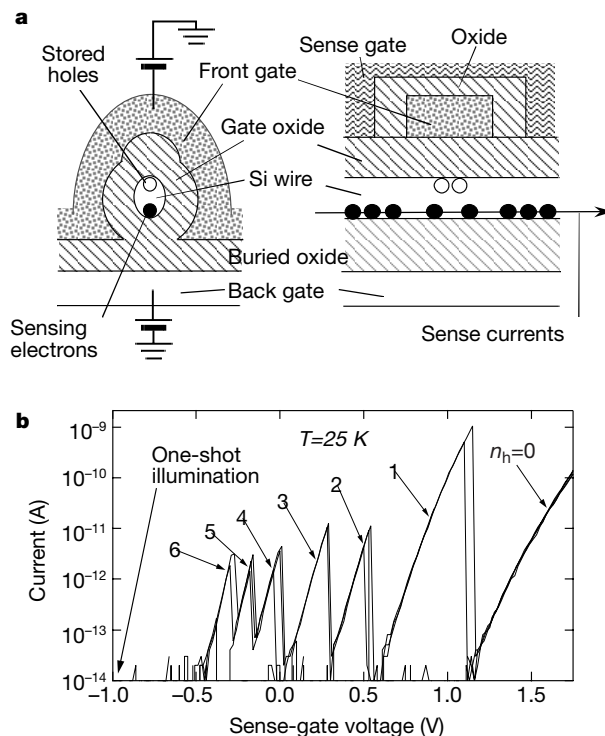


Figure 3 Our method of sensing an elementary charge. **a**, Schematic cross-sections of the Si wire, which are along the m–m' and n–n' line in Fig. 2. Electrons and holes are depicted as closed circles and open circles, respectively. A large electric field spatially separates stored holes and flowing electrons that sense the holes. **b**, Electron current versus sense-gate voltage measured at 25 K for one MOSFET. The device was illuminated once at the starting voltage and then the scan was carried out in the dark. An overlay of current curves for several runs (sequences of one-shot illumination and the voltage scan) is shown. The jump from one curve to another reflects the recombination of a stored hole. Each curve corresponds to a different number of stored holes (n_h).

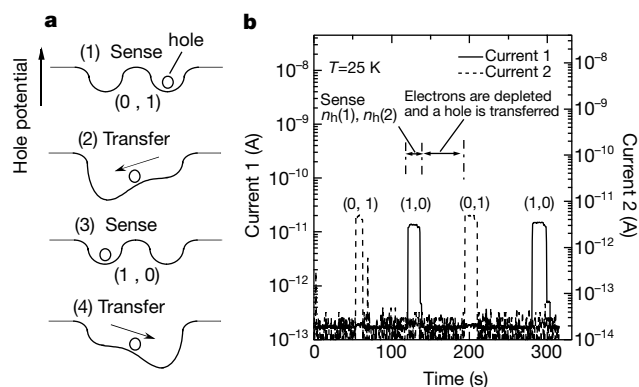


Figure 4 Demonstration of the manipulation of an elementary charge. **a**, Procedure for demonstrating CCD operation. (1) One hole is stored in MOSFET 2 and is sensed. (2) The hole is transferred to MOSFET 1. (3) The hole is sensed again. (4) The hole is transferred backwards. **b**, Results for the manipulation of a single hole between the two MOSFETs. The two drain currents are plotted for the repeated procedures of the sensing and the single-hole transfer. Owing to the sensing, the numbers of stored holes ($n_h(1)$, $n_h(2)$) are alternately detected as (0,1) and (1,0).

the Si wire to be $(1-4) \times 10^5 \text{ V cm}^{-1}$. It is large enough for electrons and holes to be quantum-mechanically confined in the triangular potential formed at the interfaces. Their ground-state wavefunction will be localized within 5 nm of the interface.

The one-shot illumination was carried out at a sense-gate voltage of -1 V . Then, the sense-gate voltage was scanned at the speed of 30 mV s^{-1} while the device was in the dark. Figure 3b shows the overlay of the results for several runs (sequences of one-shot illumination and the scan). The current rises, but suddenly drops to some lower level. This behaviour is repeated several times. The jump in current corresponds to electron-hole recombination²⁹. Because the recombination leads to a decrease in the number of the stored holes (n_h) by unity, the current curve is shifted to another position in the positive voltage direction. The recombination lifetime under the sensing was as long as a few hundred seconds, owing to the small overlap of wavefunctions between the electron and the hole. Finally, around the sense-gate voltage of 1.1 V , the current traces the curve that corresponds to the case of no stored hole. Thus we can count the exact number of holes as the curves are clearly quantized.

Figure 4a illustrates the procedure for demonstrating the single-hole transfer between the two wire-MOSFETs. In the initial state (1), one hole is stored only in MOSFET 2. To read out the number of holes in both the MOSFETs, we measure two sensing currents by setting the sense-gate voltage as high as 0.88 V . As only MOSFET 2 stores a hole, current 1 should be low and current 2 should be high. After the sensing, the sense-gate voltage is decreased to -1 V . This depletes electrons not only from the channel below the front gate but also from neighbouring channels. Next, the front-gate voltages are controlled such that the hole potential is as illustrated in (2). Then the hole can be transferred from MOSFET 2 to MOSFET 1. After that, the transferred hole is sensed again (3) and transferred backwards (4), and so on.

The results for the single-hole manipulation are shown in Fig. 4b. The two drain currents are plotted as a function of time. The sensing and the transfer of the single hole were repeated several times. To sense the hole non-destructively, the sensing time was set at a few tens of seconds that was much shorter than the electron-hole recombination lifetime. It is clear that the sensing currents reach a high level alternately. This means that the numbers of the stored holes ($n_h(1)$, $n_h(2)$) alternately change between (0,1) and (1,0), as a result of the single-hole transfer between the two wire-MOSFETs. The rather large timescale in Fig. 4b is not due to the device speed, but to the measurement system and the slow sequence of the voltage pulse. The possibility of high-frequency operation will

be investigated in future work.

Our findings are promising for the large-scale integration of single-electron transfer devices. We can design the arrangement of the MOS diodes by using branching Si wires or Si-wire networks. One possible application is the binary-decision-diagram logic circuit²⁴, where one-bit information is conveyed by the path the electron takes at the branching point of Si wires. More challenging is the application to the cellular automaton based on the single electron³⁰ or the neural network. This might be possible if the capacitive coupling between the diodes were increased.

Thus we have demonstrated the manipulation of an elementary charge in Si, using an ultrasmall CCD. By sensing an elementary charge, the transfer of a single hole between the Si wire-MOSFETs was performed at 25 K .

Received 27 November 2000; accepted 5 February 2001.

- Averin, D. V. & Likharev, K. K. in *Mesoscopic Phenomena in Solids* (eds Altshuler, B. L., Lee, P. A. & Webb, R. B.) 173–271 (Elsevier, Amsterdam, 1991).
- Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunnelling* (Plenum, New York, 1992).
- Barbe, D. F. (ed.) *Charge-Coupled Devices* (Springer, Berlin, 1980).
- Likharev, K. K. Single-electron devices and their applications. *Proc. IEEE* **87**, 606–632 (1999).
- Averin, D. V. & Likharev, K. K. Coulomb blockade of tunneling, and coherent oscillations in small tunnel junctions. *J. Low Temp. Phys.* **62**, 345–372 (1986).
- Meriav, U. & Foxman, E. B. Single-electron phenomena in semiconductors. *Semicond. Sci. Technol.* **10**, 255–284 (1995).
- Takahashi, Y. et al. Silicon single-electron devices. *Int. J. Electron.* **86**, 605–639 (1999).
- Tans, S. J. et al. Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–476 (1997).
- Likharev, K. K. Single-electron transistors: Electrostatic analogs of the DC SQUID's. *IEEE Trans. Mag.* **23**, 1142–1145 (1987).
- Fulton, T. A. & Dolan, G. J. Observation of single-electron charging effects in small tunnel junctions. *Phys. Rev. Lett.* **59**, 109–112 (1987).
- Kuzmin, L. S. & Likharev, K. K. Direct experimental observation of discrete correlated single-electron tunneling. *JETP Lett.* **45**, 495–497 (1987).
- Meriav, U., Kastner, M. A. & Wind, S. J. Single-electron charging and periodic conductance resonances in GaAs nanostructures. *Phys. Rev. Lett.* **65**, 771–774 (1990).
- Su, B., Goldman, V. J. & Cunningham, J. E. Observation of single-electron charging in double-barrier heterostructures. *Science* **255**, 313–315 (1992).
- Scott-Thomas, J. H., Field, S. B., Kastner, M. A., Smith, H. I. & Antoniadis, D. A. Conductance oscillations periodic in the density of a one-dimensional electron gas. *Phys. Rev. Lett.* **62**, 583–586 (1989).
- Ali, D. & Ahmed, H. Coulomb blockade in silicon tunnel junction device. *Appl. Phys. Lett.* **64**, 2119–2120 (1994).
- Takahashi, Y. et al. Fabrication technique for Si single-electron transistor operating at room temperature. *Electron. Lett.* **31**, 136–137 (1995).
- Dresselhaus, P. D., Ji, L., Han, S., Lukens, J. E. & Likharev, K. K. Measurement of single electron lifetimes in a multijunction trap. *Phys. Rev. Lett.* **72**, 3226–3229 (1994).
- Yano, K. et al. Room-temperature single-electron memory. *IEEE Trans. Electron Devices* **41**, 1628–1637 (1994).
- Fujiwara, A., Takahashi, Y., Murase, K. & Tabe, M. Time-resolved measurement of single-electron tunneling in a Si single-electron transistor with satellite Si islands. *Appl. Phys. Lett.* **67**, 2957–2959 (1995).
- Guo, L., Leobandung, E. & Chou, S. Y. A single-electron transistor memory operating at room temperatures. *Science* **275**, 649–651 (1997).
- Geerlings, L. J. et al. Frequency-locked turnstile device for single electrons. *Phys. Rev. Lett.* **64**, 2691–2694 (1990).
- Pothier, H. et al. Single electron pump fabricated with ultrasmall normal tunnel junctions. *Physica B* **169**, 573–574 (1991).
- Keller, M. W., Martinis, J. M., Zimmerman, N. M. & Steinbach, A. H. Accuracy of electron counting using a 7-junction electron pump. *Appl. Phys. Lett.* **69**, 1804–1806 (1996).
- Asahi, N., Akazawa, M. & Amemiya, Y. Single-electron logic device based on the binary decision diagram. *IEEE Trans. Electron Devices* **44**, 1109–1116 (1997).
- Fujiwara, A. et al. Double-island single-electron devices: A useful unit device for single-electron logic LSI's. *IEEE Trans. Electron Devices* **46**, 954–959 (1999).
- Single, C., Augke, R., Prins, F. E., Wharam, D. A. & Kern, D. P. Single-electron charging in doped silicon double dots. *Semicond. Sci. Technol.* **14**, 1165–1168 (1999).
- Cristoloveanu, S. & Li, S. S. *Electrical Characterization of Silicon-On-Insulator Materials and Devices* (Kluwer Academic, Boston, 1995).
- Coleing, J. P. *Silicon-On-Insulator Technology: Materials to VLSI* (Kluwer Academic, Boston, 1997).
- Fujiwara, A., Takahashi, Y. & Murase, K. Observation of single electron-hole recombination and photon-pumped current in Si single-electron transistors. *Phys. Rev. Lett.* **78**, 1532–1535 (1997).
- Lent, C. S., Tougaw, P. D., Porod, W. & Bernstein, G. Quantum cellular automata. *Nanotechnology* **4**, 49–67 (1993).

Acknowledgements

We are grateful to K. Murase and T. Ogino for their encouragement. We also thank T. Yamaguchi, K. Yamazaki, H. Namatsu, Y. Watanabe, J. Hayashi, T. Saito and K. Kurihara for help in the device fabrication, and S. Horiguchi, H. Inokawa and Y. Ono for valuable discussions.

Correspondence and requests for materials should be addressed to A.F. (e-mail: afuji@aetl.ntt.co.jp).

Vortex dynamics in superconducting MgB₂ and prospects for applications

Y. Bugoslavsky, G. K. Perkins, X. Qi, L. F. Cohen & A. D. Caplin

Centre for High Temperature Superconductivity, Blackett Laboratory, Imperial College, London SW7 2BZ, UK

The recently discovered¹ superconductor magnesium diboride, MgB₂, has a transition temperature, T_c , approaching 40 K, placing it intermediate between the families of low- and high-temperature superconductors. In practical applications, superconductors are permeated by quantized vortices of magnetic flux. When a supercurrent flows, there is dissipation of energy unless these vortices are 'pinned' in some way, and so inhibited from moving under the influence of the Lorentz force. Such vortex motion ultimately determines the critical current density, J_c , which the superconductor can support. Vortex behaviour has proved to be more complicated in high-temperature superconductors than in low-temperature superconductors and, although this has stimulated extensive theoretical and experimental research², it has also impeded applications. Here we describe the vortex behaviour in MgB₂, as reflected in J_c and in the vortex creep rate, S , the latter being a measure of how fast the 'persistent' supercurrents decay. Our results show that naturally occurring grain boundaries are highly transparent to supercurrents, a desirable property which contrasts with the behaviour of the high-temperature superconductors. On the other hand, we observe a steep, practically deleterious decline in J_c with increasing magnetic field, which is likely to reflect the high degree of crystalline perfection in our samples, and hence a low vortex pinning energy.

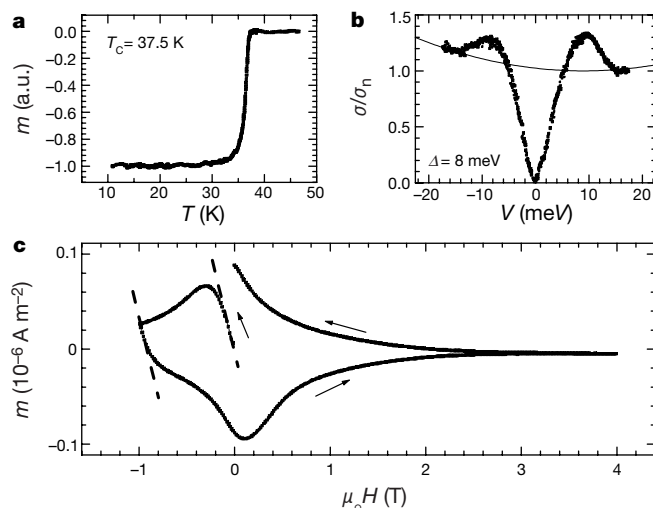


Figure 1 Characterization of superconductivity. **a**, Magnetic moment of a single small MgB₂ fragment as a function of temperature in a field of 10 mT (the sample was previously cooled in zero field), showing a sharp superconducting transition at 37.5 K. **b**, Typical tunnelling characteristic (differential conductance, $\sigma(V) = dI/dV$, normalized to its normal-state value $\sigma_n = \sigma_n(0)$) of a similar fragment; the parabola is an estimate of the background conductivity. **c**, Magnetization loop (obtained using a vibrating sample magnetometer; Model TVSM5, Oxford Instruments) at 20 K of the fragment shown in **a**; data were taken at a ramp rate of 13.3 mT s⁻¹. The initial and 'reverse leg' slopes (dashed lines) of the loop are a measure of the volume from which superconducting screening currents exclude changes in magnetic flux. If the sample is a single well-connected entity, these slopes should be equal to the actual sample volume (apart from an enhancement associated with classical demagnetizing effects).

Measurements of J_c on polycrystalline MgB₂—the only form available so far—have to be analysed carefully, because experience of the high-temperature superconductor (HTS) phases has shown that their grain boundaries impede the flow of supercurrent. This problem is particularly severe in the YBa₂Cu₃O₇ phase, which intrinsically has the highest J_c of the HTS oxides, but behaves poorly in bulk polycrystalline form, with J_c depressed by several orders of magnitude at even modest-angled grain boundaries³.

We extracted from MgB₂ powder (Alfa Aesar Co., 98% purity) several fragments of above 100 μ m size; they were somewhat irregular in shape, and X-ray diffraction indicated no obvious crystalline texture. The superconducting transition (Fig. 1a) of a single particle is sharp with a T_c of 37.5 K, and particles have well-defined tunnelling characteristics (Fig. 1b) with a gap 2Δ of $\sim 5 k_B T_c$ (for a weakly coupled BCS superconductor this factor is 3.5, for the HTS phases Bi₂Sr₂CaCu₂O₈ and YBa₂Cu₃O₇, it is 10 and 5.5, respectively^{4,5}). The key data are derived from magnetization loops that have excellent signal-to-noise ratios (Fig. 1c). Before proceeding further, we applied a number of cross-checks that have been developed in the context of HTS studies⁶. These measurements yield for this fragment a superconducting screened volume of $(7.4 \pm 0.4) \times 10^{-12}$ m³, independent of field and temperature. In a scanning electron microscope, the fragment appeared roughly ellipsoidal in shape, with cross-section axes of 450 and 250 μ m, and between 100 and 150 μ m in the third direction; its physical volume Ω is therefore $(5-7) \times 10^{-12}$ m³. The screened volume might be up to a factor of 1.5 greater than this because of the demagnetizing effects, so within the uncertainties involved, there is excellent agreement with that obtained magnetically. Thus, the straightforward conclusion is that screening currents flow in a well-connected fashion around this small polycrystalline fragment, and that over the field and temperature range studied, the grain boundaries are transparent to supercurrent, just as in low-temperature superconductor (LTS) materials. In contrast, a polycrystalline

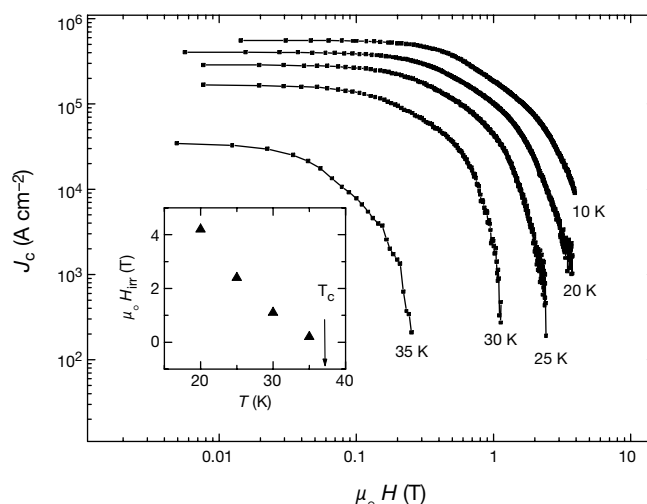


Figure 2 Critical current density J_c of a MgB₂ single fragment as a function of magnetic field. In order to convert the irreversible magnetic moment to J_c , we model the fragment as a disk of radius $a = 300$ μ m and thickness $t = 100$ μ m—that is, a volume Ω of 7×10^{-12} m³, at the upper bound of our geometrical estimate—so as to yield conservative values for J_c . The standard Bean model¹³ gives $J_c = 3\Delta m/2\Omega a$ (in SI units, but in keeping with standard practice in the field, we report J_c in units of A cm⁻² rather than A m⁻²), where Δm is the full-width of the $m(H)$ loop. Inset, the irreversibility field H_{irr} , above which irreversible (that is, screening) currents cannot be sustained, as a function of temperature; here we define H_{irr} as the field where J_c drops below the noise level, $\sim 10^3$ A cm⁻².

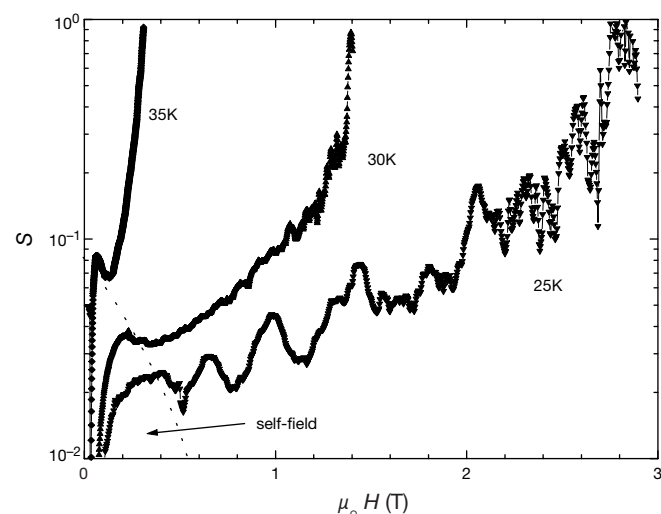


Figure 3 Vortex creep rate, S , as a function of field. We use the ‘dynamic relaxation’ method¹⁰ in which the full-width of the $m(H)$ loops depends on the field ramp rate $\dot{H}$; to a good approximation, $S = \partial \ln m / \partial \ln \dot{H}$. These data are derived from the differences of irreversible magnetic moment between loops taken at ramp rates of 13 and 3 mT s⁻¹. Because the differences are small, the signal has been increased by using a sample of ten MgB₂ fragments. The oscillations in the curves, which become more severe as the temperature is reduced, are artefacts arising from small fluctuations in sample temperature, whose effects are amplified by the differencing of the data used to extract S . In the region below the dotted line, the self-field (the field generated by the screening currents themselves) is comparable to or larger than the applied field, and so complicates the interpretation.

sample of YBa₂Cu₃O₇ (unless very well textured crystallographically) would behave as an assembly of disconnected grains at all but the lowest applied fields.

The values of J_c (Fig. 2) are high; as a benchmark, at 20 K and 1 T, J_c is 1.0×10^5 A cm⁻². However, J_c decreases quite rapidly with applied field, and an ‘irreversibility field’ H_{irr} (Fig. 2, inset) can be defined above which supercurrents cannot be sustained; that is, J_c becomes unmeasurably small. Because we have established that the sample is well connected, the J_c values shown in Fig. 2 are a reliable measure of the intragranular critical current density in MgB₂.

There have already been several magnetic studies of pressed and sintered pellets of MgB₂ (refs 7–9). All the irreversibility fields are similar to the values shown in Fig. 2, but in two cases^{7,8} the J_c values are more than an order of magnitude less than reported here. Takano *et al.*⁹, who sintered their material at high temperatures and pressures, report a rather larger J_c , $\sim 4 \times 10^4$ A cm⁻² at 20 K and 1 T, but still a factor of 2 or 3 less than we measure. Overall, these results can be understood straightforwardly: in sintered pellets of MgB₂ the intergrain contact is less perfect than in our polycrystalline as-grown fragments, to a degree dependent upon processing. The reduced cross-sectional contact area limits the magnitude of the macroscopic critical current, but its field and temperature dependences appear to be essentially independent of the source or preparation of the material—in sharp contrast to the sensitivity of the HTS phases to processing.

The other key aspect of the vortex physics is that because of thermal activation, the vortices are never totally immobile, and the screening current and associated magnetic moment decay with time in a quasi-logarithmic fashion. The relevant parameter is the vortex creep rate, S , defined as the logarithmic derivative $-\partial \ln m / \partial \ln t$. S is small at low fields (Fig. 3), but increases quasi-exponentially as the field approaches H_{irr} . The creep rate is related directly to the current–voltage (or equivalently, current density J –electric field E) characteristic of the material¹⁰, with $E \propto J^{1/5}$; thus in MgB₂ the E – J characteristics at low fields are very steep, which is desirable for

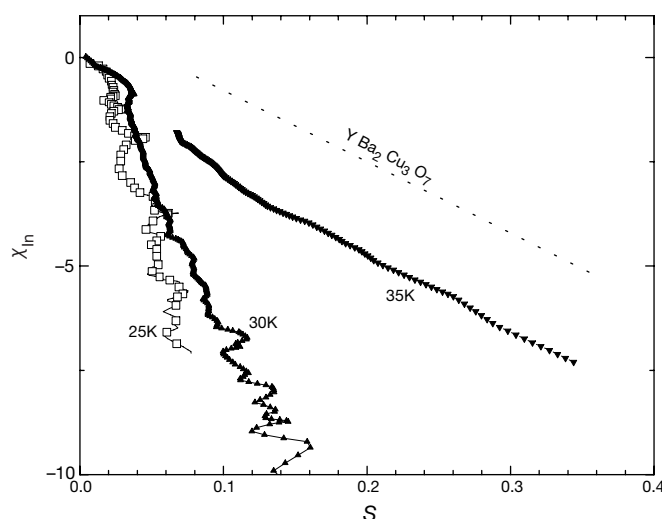


Figure 4 Comparison of the vortex behaviour in MgB₂ and YBa₂Cu₃O₇. Within the context of standard vortex pinning theory, a plot (at fixed temperature) of χ_{ln} —the dimensionless slope of the $\ln J_c$ – $\ln H$ curves of Fig. 2—against the creep rate S , with H as implicit parameter, should have some simple form¹¹. The oscillations in the data are instrumental artefacts, as described in Fig. 3 legend. The dotted line represents the behaviour of the HTS superconductor YBa₂Cu₃O₇ at temperatures not too close to T_c (ref. 11).

applications. At fields above H_{irr} the vortex response becomes linear, $E \propto J$; consequently, it is to be expected that at fields close to H_{irr} S should approach unity.

The two quantities J_c and S , measured over a range of field and temperature, together provide the essential information for an understanding of the vortex behaviour. Rather than use J_c itself, it is better to consider the dimensionless derivative $\partial \ln J_c / \partial \ln H$; that is, the slope of the curves of Fig. 2. This quantity is identical to $\partial \ln m / \partial \ln H$, and so is a logarithmic susceptibility, which we denote χ_{ln} . A plot of χ_{ln} (at a fixed temperature) against S with H as the implicit parameter has a direct physical interpretation¹¹: in YBa₂Cu₃O₇, these plots are linear, with slopes almost independent of T (except close to T_c). This linearity indicates simple power-law dependences on H of two fundamental quantities: the current density J_0 that would be attained if there were no thermal activation of the vortices, and the energy U_0 , which measures the vortex pinning as modified by vortex–vortex interactions. In MgB₂, these plots (Fig. 4) are again close to linear, and at 35 K (that is, close to T_c) have almost the same slope as that seen in YBa₂Cu₃O₇. However, as the temperature drops, they become considerably steeper, which indicates that in MgB₂ at temperatures not too close to T_c , U_0 decreases very rapidly with increasing field. Detailed measurements of χ_{ln} in conventional LTS superconductors are technically difficult, so that no comparison of this kind can be made at present.

MgB₂ is certainly a promising material in some respects— J_c is high and the grain boundaries appear benign in the context of supercurrent flow—but the irreversibility field is rather low, only about half of the upper critical field H_{c2} (refs 7, 8). The sharp rise in the vortex creep rate S as the field approaches H_{irr} , and the steep χ_{ln} – S plots, suggest that the vortex pinning interactions weaken catastrophically with increasing field. Perhaps there is a paucity of pinning defects in the ‘as-grown’ material, which would be consistent with the high degree of crystalline perfection of this compound—indicated by the very low residual resistivity¹², and by the reproducibility of T_c and H_{irr} between samples from different sources. For MgB₂ to become more useful for applications, it will be essential to find ways to enhance the vortex pinning energy. □

Received 20 February; accepted 6 March 2001.

1. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
2. Cohen, L. F. & Jensen, H. J. Open questions in the magnetic behaviour of high temperature superconductors. *Rep. Prog. Phys.* **60**, 1581–1672 (1997).
3. Mannhart, J., Chaudhari, P., Dimos, D., Tsuei, C. C. & McGuire, T. R. Critical currents in [001] grains and across their tilt boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_7$ films. *Phys. Rev. Lett.* **61**, 2476–2479 (1988).
4. Miyakawa, N. *et al.* Predominantly superconducting origin of large energy gaps in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8-x}$ from tunneling spectroscopy. *Phys. Rev. Lett.* **83**, 1018–1021 (1999).
5. Cucolo, A. M. *et al.* Quasiparticle tunneling properties of planar $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}/\text{PrBa}_2\text{Cu}_3\text{O}_{7-x}/\text{HoBa}_2\text{Cu}_3\text{O}_{7-x}$ heterostructures. *Phys. Rev. Lett.* **76**, 1920–1923 (1996).
6. Caplin, A. D., Cohen, L. F., Perkins, G. K. & Zhukov, A. A. The electric-field within high-temperature superconductors—mapping the E - J - B surface. *Supercond. Sci. Technol.* **7**, 412–422 (1994).
7. Larbalestier, D. C. *et al.* Strongly linked current flow in polycrystalline forms of the superconductor MgB_2 . *Nature* **410**, 186–189 (2001).
8. Finnemore, D. K., Ostenson, J. E., Bud'ko, S. L., Lapertot, G. & Canfield, P. C. Thermodynamic and transport properties of superconducting MgB_2 . Preprint cond-mat/0102114 at (<http://xxx.lanl.gov>) (2001).
9. Takano, Y. *et al.* Superconducting properties of MgB_2 bulk materials prepared by high pressure sintering. Preprint cond-mat/0102167 at (<http://xxx.lanl.gov>) (2001).
10. Pust, L., Jirsa, M. & Durcok, S. Correlation between magnetic hysteresis and magnetic-relaxation in YBaCuO single-crystals. *J. Low Temp. Phys.* **78**, 179–186 (1990).
11. Perkins, G. K. & Caplin, A. D. Collective pinning theory and the observed vortex dynamics in $(\text{RE})\text{Ba}_2\text{Cu}_3\text{O}_{7-x}$ crystals. *Phys. Rev. B* **54**, 12551–12556 (1996).
12. Canfield, P. C. *et al.* Superconductivity in dense MgB_2 wires. Preprint cond-mat/0102289 at (<http://xxx.lanl.gov>) (2001).
13. Bean, C. P. Magnetization of high-field superconductors. *Rev. Mod. Phys.* **36**, 31–36 (1964).

Acknowledgements

Y.B. is on leave of absence from the General Physics Institute, Moscow, Russia. We thank D. Cardwell and N. H. Babu for the supply of MgB_2 powder, and A. J. P. White for crystallographic advice. This work was supported by the UK Engineering and Physical Sciences Research Council.

Correspondence and requests for materials should be addressed to Y.B. (e-mail: y.bugoslav@ic.ac.uk).

Extreme damping in composite materials with negative-stiffness inclusions

R. S. Lakes*†‡§, T. Lee‡, A. Bersie*† & Y. C. Wang*†

* Department of Engineering Physics; † Engineering Mechanics Program; ‡ Biomedical Engineering Department; § Materials Science Program; and || Rheology Research Center, University of Wisconsin-Madison, 147 Engineering Research Building, 1500 Engineering Drive, Madison, Wisconsin 53706-1687, USA

When a force deforms an elastic object, practical experience suggests that the resulting displacement will be in the same direction as the force. This property is known as positive stiffness¹. Less familiar is the concept of negative stiffness, where the deforming force and the resulting displacement are in opposite directions. (Negative stiffness is distinct from negative Poisson's ratio^{2–6}, which refers to the occurrence of lateral expansion upon stretching an object.) Negative stiffness can occur, for example, when the deforming object has stored⁷ (or is supplied⁸ with) energy. This property is usually unstable, but it has been shown theoretically⁹ that inclusions of negative stiffness can be stabilized within a positive-stiffness matrix. Here we describe the experimental realization of this composite approach by embedding negative-stiffness inclusions of ferroelastic vanadium dioxide in a pure tin matrix. The resulting composites exhibit extreme mechanical damping and large anomalies in stiffness, as a consequence of the high local strains that result from the inclusions deforming more than the composite as a whole. Moreover, for

certain temperature ranges, the negative-stiffness inclusions are more effective than diamond inclusions for increasing the overall composite stiffness. We expect that such composites could be useful as high damping materials, as stiff structural elements or for actuator-type applications.

Composite materials were prepared with inclusions of vanadium dioxide (VO_2) in a pure (99.99%) tin matrix. Particulate inclusions 150 μm in size or smaller were incorporated into a tin matrix by rolling sheets of tin with particles, followed by casting into a cylindrical mould 3 mm in diameter. Vanadium dioxide is a ferroelastic material^{10–12} that undergoes a transformation from monoclinic to tetragonal at transformation temperature $T_c = 67^\circ\text{C}$; the transition exhibits hysteresis. It is similar in structure¹⁰ to the stiff mineral rutile (TiO_2) with elastic moduli¹³ $C_{11} = 271$ GPa and $C_{33} = 483$ GPa. By contrast, steel has a Young's modulus of $E = 200$ GPa ($C_{11} = C_{33} = 269$ GPa); for tin $E = 50$ GPa.

The material properties were measured using a broadband viscoelastic spectroscopy apparatus^{14,15}, capable of a range of eleven decades of time and frequency in torsion. Specimens were mounted using cyanoacrylate cement: at the top to its support rod, and at the bottom, to a high magnetic intensity neodymium iron boron magnet. The dynamic experiments were conducted by applying a sinusoidal voltage at 100 Hz, from a digital function generator to a Helmholtz coil. The lowest specimen resonance was at several kilohertz. The coil imposed a magnetic field on the permanent magnet and transmitted an axial torque to the specimen. The angular displacement of the specimen was measured using laser light reflected from a mirror mounted on the magnet to a split-diode light detector. The detector signal was amplified with a wide-band differential amplifier. Torque was inferred from the Helmholtz coil current, supported by calibrations using the well characterized 6061 Al alloy²⁹. Input and output voltages and the phase between them were recorded using a lock-in amplifier (SRS 850, Stanford Research Systems). Data were reduced using the analytical relationship for the torsional rigidity of a viscoelastic cylinder. Constituent stiffness was tuned by varying the temperature. An insulated inner chamber was heated to above 85°C and the temperature was maintained for at least 30 minutes. The heater was shut off and the chamber allowed to cool. Cooling rate over the temperature range of interest was less than 1°C per minute.

Material properties, stiffness and damping ($\tan\delta$, where δ is the phase between stress and strain) as a function of temperature of a particulate VO_2 -tin composite with 1% by volume of inclusions are shown in Fig. 1. Although the concentration of inclusions is dilute, large anomalies are observed in the mechanical damping, $\tan\delta$, as well as in the stiffness. For comparison, $\tan\delta$ of pure Sn is 0.019 and varies little with temperature, as shown. The sharp dependence on temperature we attribute to the fact that the inclusions are much stiffer than the matrix (away from the transition temperature) and can only balance the matrix stiffness near the transition. The peaks observed in the present 1% VO_2 composite are larger than in 100% VO_2 , which had a peak¹² of 0.04 to 0.05 in $\tan\delta$ and a dip of 8% in stiffness at the transition. In the composite, there is interplay between constituents of positive and negative stiffness; by contrast, 100% VO_2 has positive stiffness at all temperatures due to the formation of domains.

A negative-stiffness lumped element (a structure) can be stabilized by a hard constraint. For an unconstrained block (or, surface traction boundary condition, in the language of elasticity¹) of isotropic material to be stable, the shear modulus G and the bulk modulus B , as stiffness quantities, must be positive. This corresponds to a range of Poisson's ratio ν (defined as the negative transverse strain of a stretched or compressed body divided by its longitudinal of strain) $-1 < \nu < 0.5$. Materials of negative Poisson's ratio, while exhibiting the unusual property of transverse expansion when stretched, can have all positive-stiffness values, and hence stability. For most solids¹, ν is between 0.25 and 0.33.

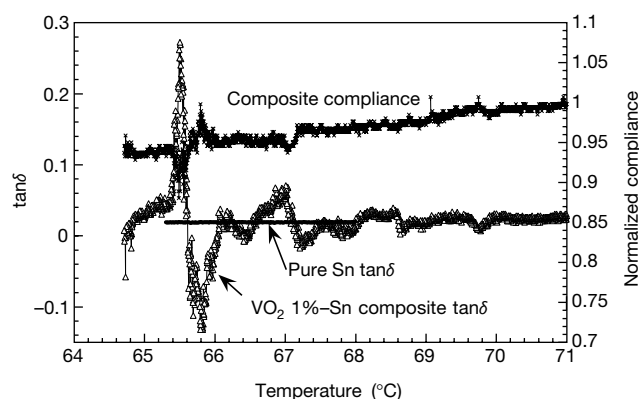


Figure 1 Experimental torsional compliance (inverse stiffness) and mechanical damping, $\tan\delta$, versus temperature. Open triangles show damping of a composite containing 1% by volume vanadium dioxide particles in a tin matrix. Crossed points show damping of pure tin for which $\tan\delta = 0.019$ over the temperature range considered. Measurements were conducted at 100 Hz, well below resonance, during slow cooling through the ferroelastic transition of the inclusions.

Foams^{2,3} with $\nu \approx -0.7$, and polycrystals⁴, crystals⁶, and other microstructures⁵ with a negative Poisson's ratio have been produced; they are stable. For a constrained isotropic block (surface displacement boundary condition¹⁶), stability requires $G > 0$ and $-\infty < \nu < 0.5$ or $1 < \nu < \infty$. As Young's modulus is $E = 2G(1 + \nu)$, an isotropic material with Young's modulus less than zero is stable provided it is constrained. Another criterion of stability is strong ellipticity¹⁷, for which $G > 0$ and $\nu < 0.5$ or $\nu > 1$, allowing negative Young's and bulk moduli. Violation of strong ellipticity gives rise to instability: bands of heterogeneous deformation¹⁷ form in materials. In single domains of ferroelastic material, band formation is suppressed by surface energy considerations. Therefore single domain particles, when constrained by a sufficiently stiff composite matrix, can be stable. So inclusions can have negative Young's modulus, and if they are small enough, even negative shear modulus.

The stabilizing constraint need not be perfectly rigid; it can be the matrix phase of a composite. Specifically, we have shown⁹ via composite theory that negative-stiffness elastic materials can be stabilized by embedding them as polydisperse spherical inclusions in an isotropic composite governed by the Hashin–Shtrikman shear modulus (G_L) formula¹⁸

$$G_L = G_2 + \frac{V_1}{\frac{1}{G_1 - G_2} + \frac{6(K_2 + 2G_2)V_2}{5(3K_2 + 4G_2)G_2}} \quad (1)$$

where K_1 and K_2 , G_1 and G_2 , and V_1 and V_2 are the bulk modulus, shear modulus and volume fraction of phases 1 and 2, respectively. Although we can show that a system is stable with respect to certain modes of instability, we cannot confidently prove that a complex system is stable with respect to all possible modes, of which there may be an infinite number. Therefore experiments are necessary.

For viscoelastic materials, the moduli become complex following the elastic–viscoelastic correspondence principle^{19,20}. When inclusion phase 1 has a negative shear modulus of about -1.1 of the matrix modulus, composite behaviour²¹ is predicted to exhibit a large peak in damping, $\tan\delta$, and an anomaly in stiffness. As an illustration of the concept, Fig. 2 shows representative behaviour versus inclusion stiffness based on composite theory with isotropic inclusions and matrix. The abscissa is normalized inclusion stiffness, which is tuned indirectly by temperature in the experiments. If inclusion concentration is increased or matrix damping is reduced, theory allows the composite to have negative stiffness and $\tan\delta$. The

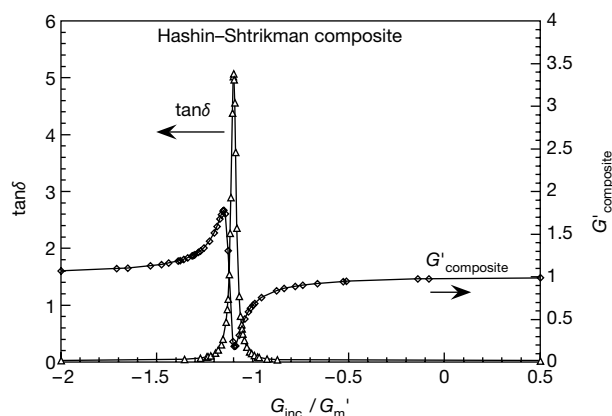


Figure 2 Theoretical shear modulus G' (real part) and mechanical damping, $\tan\delta$, versus inclusion shear modulus G_{inc} normalized to matrix shear modulus G'_m , of Hashin–Shtrikman composite containing 1% of spherical particulate inclusions by volume. Inclusions are allowed to have a negative shear modulus. Matrix damping is assumed to be 0.025. Inclusion damping is assumed to be zero. Similar behaviour is predicted in the bulk properties if the inclusions have a negative bulk modulus.

theory is simplistic in that inclusion anisotropy and heterogeneity of their environment is ignored.

Indirect experimental evidence of negative shear modulus is found in the banding instability which occurs in compressed foams²². In ferroelastic, ferroelectric and ferromagnetic materials²³ the bands, called domains, occur below a transformation temperature T_c . Small particles may be single domains²³, owing to surface energy; domain size can be from micrometres to centimetres depending on the material. A negative-stiffness material may be represented by an energy versus deformation diagram in which there are at least two relative minima or energy wells. Negative stiffness was directly observed experimentally in tetrakaidecahedral single-cell models of foam materials, in compression under displacement control constraint²². The concept of composites with negative stiffness is supported by experimental study of a constrained, lumped one-dimensional model system involving a single buckled tube as a negative stiffness element²⁴.

Observed stiffness and damping anomalies in the composite are much larger than they could be for inclusions of any positive stiffness or damping. For example, if the particles were as stiff as diamond ($E = 1,000$ GPa), composite theory¹⁸ predicts a composite stiffening effect of only 1.9% and no change in $\tan\delta$; if the particles were infinitely stiff, the composite would be only 2.1% stiffer than a tin matrix. If the particle stiffness were to vanish, the composite would soften by 1.9% and have no change in $\tan\delta$. The inclusions are therefore more effective than diamond in increasing the composite stiffness at selected temperatures; moreover, the composite exceeds the classical bounds¹⁸ based on positive stiffness. The observed peaks in damping cannot be explained by any variation in the properties of positive-stiffness inclusions. Even if inclusion $\tan\delta$ were as large as 1.0 at the transition, composite damping could increase from 0.025 to at most 0.032, assuming positive inclusion stiffness. Therefore the inclusions exhibit negative stiffness.

Composites with inclusions of negative stiffness may be called extenlibral because they are on the boundary of balance, or archi-dynamic because they are based on initial force. They are pertinent to any heterogeneous material in which one constituent undergoes a phase transformation and another does not; also to some materials with a pre-strained constituent. They may be of use in studying the properties of single domains of ferroelastic, ferroelectric, ferromagnetic, or shape memory martensite²⁵ materials. Because a dilute concentration is sufficient to obtain substantial effects, not much sample material is needed. Some materials can be easily prepared as

large single crystals; polycrystalline arrays may be brittle. Such composites may be of use as high-performance damping materials, since the figure of merit²⁶ in the present results exceeds that of commonly used materials ($E \tan \delta \approx 0.6$ GPa) by a factor of more than 20. For that purpose, sensitivity to temperature could be reduced via matching inclusion and matrix stiffness. Moreover, pre-stressed or pre-buckled²⁴ elements with minimal temperature sensitivity may be used as inclusions. Bounds on properties of complex heterogeneous materials are generally derived assuming positive phase properties¹⁸. These bounds can be exceeded if negative stiffness is allowed, permitting extreme properties not previously anticipated. As in thermoelastic and piezoelectric materials, elasticity is coupled with temperature and electric field respectively, these composites may find use in high-performance sensors and actuators. These composites may also occur naturally in rocks and in biological materials (in which anomalies were reported²⁷); they may be considered in the context of deep-focus earthquakes and attenuation of seismic waves. □

Received 25 July 2000; accepted 2 February 2001.

1. Timoshenko, S. P. & Goodier, J. N. *Theory of Elasticity* 3rd edn (McGraw-Hill, New York, 1970).
2. Lakes, R. S. Foam structures with a negative Poisson's ratio. *Science* **235**, 1038–1040 (1987).
3. Lakes, R. S. Advances in negative Poisson's ratio materials. *Adv. Mater.* **5**, 293–296 (1993).
4. Haeri, A. Y., Weidner, D. J. & Parise, J. B. Elasticity of α -cristobalite: a silicon dioxide with a negative Poisson's ratio. *Science* **257**, 650–652 (1992).
5. Rothenburg, L., Berlin, A. A. & Bathurst, R. J. Microstructure of isotropic materials with negative Poisson's ratio. *Nature* **354**, 470–472 (1991).
6. Baughman, R. H., Shacklette, J. M., Zakhidov, A. A. & Stafstrom, S. Negative Poisson's ratios as a common feature of cubic metals. *Nature* **392**, 362–365 (1998).
7. Thompson, J. M. T. Stability prediction through a succession of folds. *Phil. Trans. R. Soc. Lond.* **292**, 1–23 (1979).
8. Thompson, J. M. T. 'Paradoxical' mechanics under fluid flow. *Nature* **296**, 135–137 (1982).
9. Lakes, R. S. & Drugan, W. J. Stiff elastic composite materials with a negative stiffness phase. *J. Mech. Phys. Solids* (submitted).
10. Heckingbottom, R. & Linnett, J. W. Structure of vanadium dioxide. *Nature* **194**, 678 (1962).
11. Paquet, D. & Leroux-Hagon, P. Electron correlations and electron-lattice interactions in the metal-insulator ferroelastic transition in VO₂: a thermodynamical study. *Phys. Rev. B* **22**, 5284–5301 (1979).
12. Zhang, J. X., Yang, Z. H. & Fung, P. C. W. Dissipation function of the first-order phase transformation in VO₂ ceramics by internal friction measurements. *Phys. Rev. B* **52**, 278–284 (1995).
13. *Handbook of Chemistry and Physics* C-40 12–40 (CRC Press, Boca Raton, Florida, 1996).
14. Brodt, M., Cook, L. S. & Lakes, R. S. Apparatus for measuring viscoelastic properties over ten decades: refinements. *Rev. Sci. Instrum.* **66**, 5292–5297 (1995).
15. Lakes, R. S. & Quackenbush, J. Viscoelastic behaviour in indium tin alloys over a wide range of frequency and time. *Phil. Mag. Lett.* **74**, 227–232 (1996).
16. Bramble, J. H. & Payne, L. E. On the uniqueness problem in the second boundary value problem in elasticity. *Proc. 4th Natl Cong. Appl. Mech.* 469–473 (American Society of Mechanical Engineers, Berkeley, California, 1963).
17. Knowles, J. K. & Sternberg, E. On the failure of ellipticity and the emergence of discontinuous gradients in plane finite elastostatics. *J. Elasticity* **8**, 329–379 (1978).
18. Hashin, Z. & Shtrikman, S. A variational approach to the theory of the elastic behavior of multiphase materials. *J. Mech. Phys. Solids* **11**, 127–140 (1963).
19. Read, W. T. Stress analysis for compressible viscoelastic materials. *J. Appl. Phys.* **21**, 671–674 (1950).
20. Hashin, Z. Viscoelastic behavior of heterogeneous media. *J. Appl. Mech. Trans. ASME* **32E**, 630–636 (1965).
21. Lakes, R. S. Extreme damping in composite materials with a negative stiffness phase. *Phys. Rev. Lett.* (in the press).
22. Rosakis, P., Ruina, A. & Lakes, R. S. Microbuckling instability in elastomeric cellular solids. *J. Mater. Sci.* **28**, 4667–4672 (1993).
23. Salje, E. *Phase Transitions in Ferroelastic and Co-elastic Crystals* 10, 72 (Cambridge Univ. Press, Cambridge, 1990).
24. Lakes, R. S. Extreme damping in compliant composites with a negative stiffness phase. *Phil. Mag. Lett.* **81**, 95–100 (2001).
25. Ren, X. & Otsuka, K. Origin of rubber-like behaviour in metal alloys. *Nature* **389**, 579–583 (1997).
26. Cremer, L., Heckl, M. A. & Ungar, E. E. *Structure Borne Sound* 2nd edn 243–247 (Springer, Berlin, 1988).
27. Pugh, J. W., Rose, R. M., Paul, I. L. & Radin, E. L. Mechanical resonance spectra in human cancellous bone. *Science* **181**, 271–272 (1973).
28. Falk, F. Model free energy, mechanics, and thermodynamics of shape memory alloys. *Acta Metall.* **28**, 1773–1780 (1980).
29. Duffy, W. Acoustic quality factor of aluminum alloys from 50 mK to 300 K. *J. Appl. Phys.* **68**, 5601–5609 (1990).

Acknowledgements

We thank W. Drugan and R. Cooper for supportive comments and discussions. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to R.S.L. (e-mail: lakes@engr.wisc.edu).

Climate variability 50,000 years ago in mid-latitude Chile as reconstructed from tree rings

Fidel A. Roig^{*}, Carlos Le-Quesne^{†‡}, José A. Boninsegna^{*}, Keith R. Briffa[§], Antonio Lara[‡], Håkan Grudd^{||}, Philip D. Jones[§] & Carolina Villagrán[¶]

^{*} Laboratorio de Dendrocronología, IANIGLA-CONICET, CC 330 (5500) Mendoza, Argentina

[†] Department Botánica, Facultad de Biología, Universidad de Oviedo, Catedrático Rodrigo Uría s/n, 33006, Spain

[‡] Instituto de Silvicultura, Universidad Austral de Valdivia, Casilla 567, Valdivia, Chile

[§] Climate Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

^{||} Department of Physical Geography, Stockholm University, S-10691 Stockholm, Sweden

[¶] Department Biología, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

High-resolution proxies of past climate are essential for a better understanding of the climate system¹. Tree rings are routinely used to reconstruct Holocene climate variations at high temporal resolution², but only rarely have they offered insight into climate variability during earlier periods³. *Fitzroya cupressoides*—a South American conifer which attains ages up to 3,600 years—has been shown to record summer temperatures in northern Patagonia during the past few millennia⁴. Here we report a floating 1,229-year chronology developed from subfossil stumps of *F. cupressoides* in southern Chile that dates back to approximately 50,000 ¹⁴C years before present. We use this chronology to calculate the spectral characteristics of climate variability in this time, which was probably an interstadial (relatively warm) period. Growth oscillations at periods of 150–250, 87–94, 45.5, 24.1, 17.8, 9.3 and 2.7–5.3 years are identified in the annual subfossil record. A comparison with the power spectra of chronologies derived from living *F. cupressoides* trees shows strong similarities with the 50,000-year-old chronology, indicating that similar growth forcing factors operated in this glacial interstadial phase as in the current interglacial conditions.

With the exception of polar and tropical ice cores and some ocean-sediment cores, palaeoclimate records with annual resolution based on studies of tree rings, coral, and varved sediments have been recovered only for the late Holocene⁵. Here we report annually resolved proxies of palaeoclimate data for southern South America for an interstadial period thought to correspond to marine oxygen isotope stage 3 (OIS 3; ref. 6). Our analysis is based on subfossil tree remnants found at Seno Reloncaví in the southern Lake District of Chile (40° 00' to 42° 30' S, 71° 30' to 74° 00' W). Here we use the term 'subfossil' to mean wood preserved since the late Pleistocene epoch that still contains carbon material. Supported by glacial geology and palynology studies, the southern Lake District represents an area where late Quaternary climate fluctuations have been best established for the entire Andes^{7–10}. Although the Lake District was affected repeatedly by large ice lobes descending from the Andes at times > 14 kyr before present (BP), long-term full-glacial stratigraphic pollen records show that the forests did not completely disappear during the periods of ice advance and rapidly re-expanded afterwards^{8,9}. After ice recession, evergreen forest, mainly dominated by broad-leaved and conifer elements, expanded across the Lake District with a wide ecological amplitude, from lowland floodplains to near the modern, upper tree line.

Annually resolved records in the region based on the characteristic

tree of the north Patagonian evergreen forest, *F. cupressoides* (alerce)—which attains ages of 3,600 years¹¹—were available previously only for the past few millennia. *Fitzroya*, a monotypic conifer restricted to wet environments, is a sensitive indicator of climate; both the width and the density of the annual rings respond primarily to variations of summer (December to March) temperatures^{11–13}. In southern Chile, present seasonal climates are regulated by the annual cycle of latitudinal migration of the South Pacific Subtropical High coupled with the cold off-shore Humboldt ocean current and the orographic effects of the Andes¹⁴. The prevailing westerly storm tracks are related to equatorward shifts of the polar front. These low-pressure storms, which are most frequent during winter, are responsible for the humid and equable climate that supports the temperate rain forests.

In the forested landscape of southern Chile and Argentina, acidic soils and climate are conducive to the preservation of tree remnants suitable for radiocarbon dating and dendrochronological research^{6,15}. After the 1960 earthquake¹⁶ that affected the southern Lake District, an intertidal area alongside the north shore of Seno Reloncaví became subject to marine erosion. Well preserved subfossil *F. cupressoides* stumps were exposed; the trees had been originally buried *in situ* by a lahar¹⁷ (a debris flow composed chiefly of volcanoclastic materials). Particularly at the Pelluco site (6 km east of Puerto Montt, 41.5° S, 72.9° W), these stumps attain diameters of up to 1–1.5 m. Few stumps recovered from Pelluco belong to *Pilgerodendron uviferum*. This is evocative of the present-day *Fitzroya*–*Pilgerodendron* community on marshy, low-drainage soils¹³. Because the outermost rings have been eroded from most of the cross-sections, it was impossible to ascertain if the trees were killed and buried simultaneously in sediment. However, the position and orientation of the stumps suggest that this forest was affected by a catastrophic single event. Tephra deposition, flooding and abrupt changes of the ground level—caused, for example, by earthquakes—today affect the stability of *F. cupressoides* forests¹⁸; these factors must have operated similarly in the past.

The AMS (accelerator mass spectrometry) ¹⁴C ages of three wood samples containing the five innermost rings are 49,780 ± 2,040 ¹⁴Cyr BP, 49,770 ± 2,040 ¹⁴Cyr BP, and 49,370 ± 2,000 ¹⁴Cyr BP (CAMS 46534, 46535, and 46536 respectively). Previous radiocarbon dating showed similar results^{17,19}. Because these dates are close to or higher than the limit for ¹⁴C dating, they should be

regarded as minimum ages for the subfossil wood. According to the chronological and stratigraphic glacial record proposed for the area⁷, the wood samples pre-date the Last Glacial Maximum, and broadly correspond to the early portions of middle Llanquihue time (correlated with the Middle Wisconsinan/Weichselian glacial stages in the Northern Hemisphere). Furthermore, the longer pollen records available for the region^{8,9} document the existence of an arboreal taxa association with a significant presence of *Fitzroya*-related pollen types (*Pilgerodendron* type) in the early phases of middle Llanquihue time (that is, about 47,000 ¹⁴Cyr BP). This implies that tracts of interstadial forest immediately adapted to growth under the cool temperate, humid, climate regime in the mid-latitudes of southern Chile at the early stages of OIS 3. The simplest explanation, pending more precise dates, is that the subfossil *Fitzroya* stand became established during an interval of humid and cool climate, broadly corresponding to the onset of OIS 3.

Subfossil wood samples from Pelluco were collected by different workers from Chile, Sweden and Argentina, and processed according to standard dendrochronological techniques²⁰. 51 polished transverse cross-sectional surfaces from 30 trees were measured (with a precision of ± 0.01 mm), and the resulting individual ring-width series were subjected to extensive cross-dating comparisons to identify missing rings and eliminate other possible errors at the measuring stage. After this procedure, 47 radii from 28 trees were selected and used to construct a floating mean tree-ring chronology 1,229 years long (the range of specimen length was 185–1,154 years). Tree-ring standardization²⁰ was performed to remove the inherent long-term growth trend associated with tree age and forest dynamics. Individual series were detrended using a smoothing spline with a fixed 50% frequency response of 200 years, resulting in dimensionless indices, before averaging them into the chronology. This standardization procedure preserves most of the century-timescale tree growth variations²¹. The final chronology, with the corresponding internal sample replication, is shown in Fig. 1. To our knowledge, this is the oldest known floating (that is, not absolutely dated) tree-ring chronology developed to date.

To identify cycles contained in our Pelluco chronology, spectral analysis was performed using two different methods as defined in Fig. 2. Fourier analysis (using the Blackman-Tukey technique) showed that long-term cyclic components exceeding the *a priori*

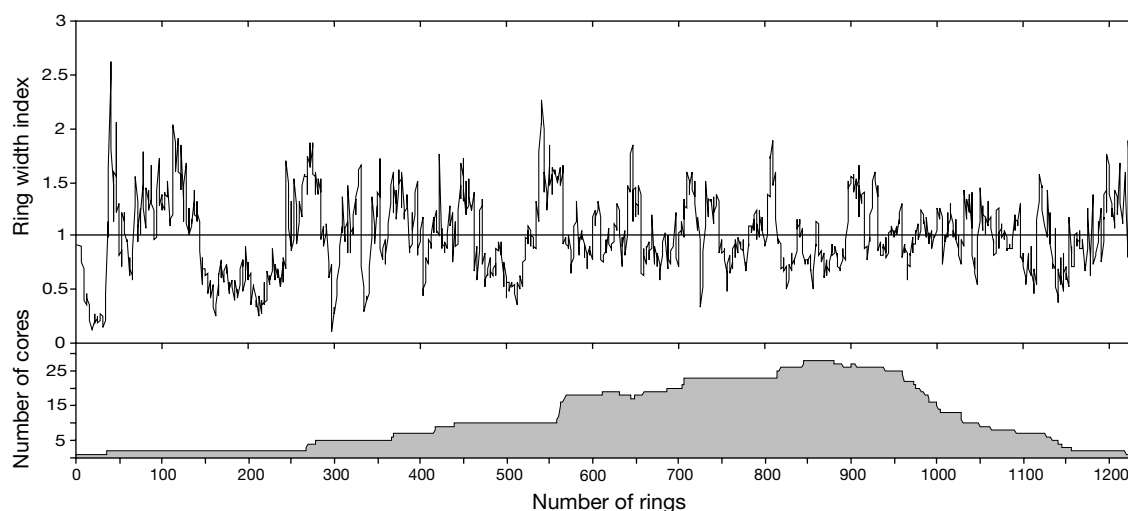


Figure 1 The 1,229-year-long floating tree-ring chronology from Pelluco, southern Chile. A cross-dated set of 47 ring-width radii obtained from 28 subfossil trees make up the ARSTAN (AutoRegressive STANdardization) chronology, whose best replicated section (more than 5 series) exists between relative rings 278 and 1,139. The subfossil chronology shows notable year-by-year changes in ring widths. Those periods with very narrow rings (two cells wide) make measurement and cross-dating rather difficult.

Missing rings were infrequent. The overall series mean ring-width is 0.42 mm and the cross-correlation average between all series is 0.53, with a mean series length of 430 years. The signal-to-noise ratio is relatively high at 5.25, and the variance explained by the first eigenvector is 32.27%. The first-order autocorrelation is 0.85 and the mean sensitivity is 0.15. These values are similar to the statistics of living *Fitzroya* chronologies^{4,27}.

95% significance level occur at periods of 153–245 and 87–94 years, which respectively account for 21.38% and 19.71% of the variance in the power spectrum. Other significant peaks in the spectrum, which account for the 13% of total variance, fall within the higher frequency band (periodicities in the range of 2–7 years; Fig. 2a inset). The same cycles are reproduced in the maximum entropy power spectrum (not shown here).

The Blackman-Tukey and maximum-entropy spectral properties of the chronologies derived from living *Fitzroya* trees are broadly

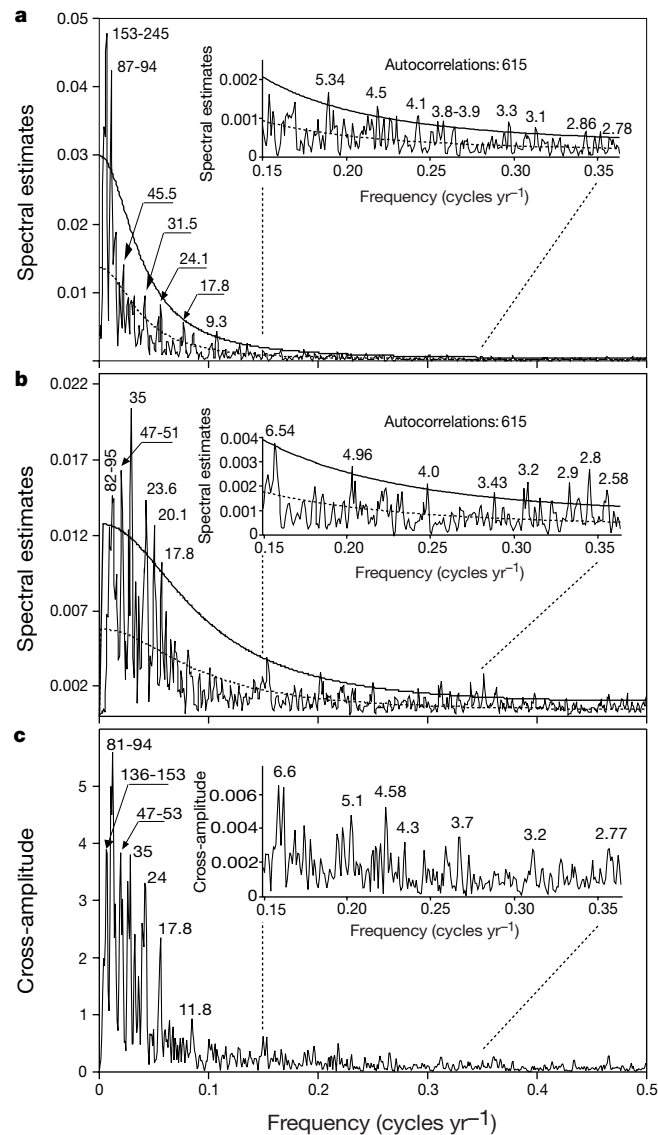


Figure 2 The spectral and cross-spectral characteristics of the subfossil (a) and living (b) *F. cupressoides* woods. We determined that 615 lags of the autocorrelation function (~50% the series length) were adequate to resolve the frequency spectrum with high resolution and moderate stability. The 95% confidence limits of the Blackman-Tukey²⁸ (BT) spectrum were estimated from a “red noise” null continuum based on the lag-1 autocorrelation coefficient. The periods are given in years for each individual peak. The (Hamming) bandwidth that we used was equal to 5. The maximum entropy²⁹ (ME) spectrum (not shown) was estimated from a 123-term prediction error filter (10% of the series length) using Burg’s³⁰ algorithm. For comparison, the same analyses were performed for a similar time span (1,229-years) on a set of three millennium-long living *Fitzroya* chronologies (Lenca¹¹, Cisne²⁷ and Costa³¹ chronologies, respectively) standardized in a similar way as the subfossil chronology. In this case we choose the amplitudes of the first eigenvector to develop our spectral estimates. The spectra are shown over the frequency range 0–0.50. c, The cross-spectral relationships between subfossil and modern chronologies. Expanded details of each power spectrum are shown in the insets.

similar to those identified in the subfossil series (that is, the peaks at 82–95, 23.6 and 17.8 years in the living trees are close to peaks at 87–94, 24.1 and 17.8 years detected in the subfossil chronology; Fig. 2a, b). In the low-frequency band of 80–254 years, the subfossil chronology shows a high concentration of variance (41.09%) but only 12.49% is reached in the modern chronology. Peaks at 47–51, 35, 23.6, 20.1 and 17.8 years are statistically significant in the modern chronology but do not exceed the *a priori* 95% confidence level in the Blackman-Tukey spectrum from the subfossil series. Association between the spectra of the living and subfossil records is also observed through a cross-spectra comparison in the frequency domain (Fig. 2c). The variances in the cross-amplitude spectra show relatively strong correspondence over the resolved frequency range equivalent to periods centred at 136–153, 81–94, 47–53, 35, 24, 17.8 years and other peaks over the higher-frequency range between 6.6 and 2.14 years (Fig. 2c inset). The broad correspondence between spectra suggest the possibility that similar processes have influenced the radial growth of both subfossil and modern *Fitzroya* trees.

Focusing on longer timescales, it is evident from the spectral density distribution that the ~150–250-year waveform is a characteristic of the Pelluco chronology, and may reflect low-frequency variations of climate during the late Pleistocene. Oscillations of the order of ~200 years have been found in late Pleistocene–Holocene records from marine cores from the west side of the Antarctic peninsula²². Solar modulation may be a possible forcing mechanism at this frequency because there is evidence of a near 220-year ¹⁴C cycle as a result of sunspots²³. Other cyclic fluctuations in the Pelluco chronology on a decadal to centennial timescale may also be attributed to varying solar activity. Atmospheric $\Delta^{14}\text{C}$ oscillations derived from tree-ring chronologies covering almost the entire Holocene (a 9,600-yr record)²³ show changes with periodicities close to 22 and 90 years, cycles related respectively to Hale and Gleissberg global solar changes. In a similar way, the oscillatory modes detected in the subfossil tree-ring chronology at the 87–94-year and the 24.1-year periods may correspond to periodicities related to both Hale and Gleissberg solar cycles contained in the ‘modern’ (Holocene) tree-ring records.

Determining the forcing mechanisms that produced the short-term cycles in our record is less straightforward. The El Niño/Southern Oscillation (ENSO) appears to be the largest single source of present-day global inter-annual climate variability. ENSO has operated for most of the Holocene, and strong climate shifts associated with El Niño events have been documented by, among other proxies, South American tree rings^{4,12,13}. But for earlier times, data from marine cores imply that during glacial sea-level low-stands, the ocean/atmosphere conditions in the inter-tropical Pacific prevented or dramatically reduced the functioning of the ENSO system^{24,25}. Furthermore, there are indications that the ENSO system probably operated during middle–late Pleistocene interglacials and warm interstadial periods, although overall climate and boundary conditions at these stages may have been too different to permit the characteristic ENSO climate variability in terms of frequency, intensity and recurrence²⁶. Thus, some of the high-frequency oscillations at timescales of 2–7 years observed in our subfossil chronology still remain to be assessed in relation to other ancient high-resolution records.

F. cupressoides tree-rings are known to correlate well with summer temperatures^{4,11,13}. This relationship permits reconstruction of Late Holocene climate fluctuations in the northern Patagonian Andes, at least during the past two millennia⁴. Uniformitarianism suggests that most of the tree-growth oscillations observed in our subfossil tree-ring Pelluco chronology may be attributed to summer temperature variability during Pleistocene interstadials at early stages of OIS 3. Hence, the oscillations of the order of 150–250 years detected in this chronology may have the same forcing as century-scale oscillations observed in temperature reconstructions of the Late

Holocene derived from tree-ring data⁴.

Our study suggests that comparable cycles in tree growth occurred between interstadials of the last glaciation and today, and hence that similar factors have affected the radial growth of *Fitzroya* since the Late Pleistocene. Considering that the Earth's atmosphere—and the ocean/atmosphere system—were significantly disturbed during the glacial cycles of the Late Pleistocene, the notable similarity in the spectral properties of the Pelluco subfossil and modern *F. cupressoides* chronologies suggests that the forcing mechanisms of climate during the interstadials have not changed dramatically. The potential for developing very long tree-ring chronologies—from the present day back several millennia—is one of the most promising results in southern South America dendrochronology^{15,27}. Our Pelluco tree-ring chronology represents a high-resolution window that allows us the possibility of analysing fluctuations in tree growth and climate during warmer events of the last glacial stage in southern Chile, where other very old palaeorecords (>50,000 yr) are rare. □

Received 17 October 2000; accepted 18 January 2001.

1. Oldfield, F. (ed.) Global Change Report No. 45 (International Geosphere-Biosphere Programme (IGBP), Stockholm, 1998).
2. Fritts, H. C. *Reconstructing Large-Scale Climatic Patterns from Tree-Ring Data* (Univ. Arizona Press, Tucson, 1991).
3. Pilcher, J. R., Baillie, M. G. L., Schmidt, B. & Becker, B. A 7,272-year tree-ring chronology for western Europe. *Nature* **312**, 150–152 (1984).
4. Villalba, R. et al. in *Climate Fluctuations and Forcing Mechanisms of the Last 2,000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) 161–189 (Springer, Berlin, 1996).
5. Bradley, R. S. *Quaternary Paleoclimatology* (Chapman & Hall, London, 1992).
6. Clapperton, C. *Quaternary Geology and Geomorphology of South America* (Elsevier, Amsterdam, 1993).
7. Denton, G. H. et al. Geomorphology, stratigraphy, and radiocarbon chronology of Llanquihue Drift in the area of the southern Lake District, Seno Reloncavi, and Isla Grande de Chiloé, Chile. *Geografis. Ann.* **81**, 167–229 (1999).
8. Heusser, C. J., Heusser, L. E. & Lowell, T. V. Paleocology of the southern Chilean Lake District—Isla Grande de Chiloé during middle—late Llanquihue glaciation and deglaciation. *Geografis. Ann.* **81**, 231–284 (1999).
9. Heusser, C. J. et al. Pollen sequence from the Chilean Lake District during the Llanquihue glaciation in marine oxygen Isotope Stages 4–2. *J. Quat. Sci.* **15**, 115–125 (2000).
10. Denton, G. H. et al. Interhemispheric linkage of paleoclimate during the last glaciation. *Geografis. Ann.* **81**, 107–153 (1999).
11. Lara, A. & Villalba, R. A 3620-year temperature record from *Fitzroya cupressoides* tree rings in southern South America. *Science* **260**, 1104–1106 (1993).
12. Villalba, R. Tree-ring and glacial evidence for the Medieval Warm Epoch and the Little Ice Age in southern South America. *Clim. Change* **26**, 183–197 (1994).
13. Roig, F. A. Dendroklimatologische Untersuchungen an *Fitzroya cupressoides* im Gebiet der Küstenkordillere und der Südlichen Anden. Thesis, Basel Univ. (1996).
14. Miller, A. in *Climates of Central and South America* (ed. Schwerdtfeger, W.) 113–145 (Elsevier, Amsterdam, 1976).
15. Roig, F. A., Roig, C., Rabassa, J. & Boninsegna, J. A. Fuegian floating tree-ring chronologies from sub-fossil *Nothofagus* woods. *Holocene* **6**, 469–476 (1996).
16. Lomnitz, C. Major earthquakes and tsunamis in Chile during the period 1535 to 1965. *Geol. Rundsch.* **59**, 938–960 (1970).
17. Klohn, C. Beobachtungen über die Reste eines späteiszeitlichen Alcereswaldes. *Z. Naturfreunde Wanderer* **1975–1976**, 75–78 (1976).
18. Lara, A. The dynamics and disturbance regimes of *Fitzroya cupressoides* forests in the south-central Andes of Chile. Thesis, Univ. Colorado (1991).
19. Porter, S. C. Pleistocene glaciation in the southern Lake District of Chile. *Quat. Res.* **16**, 263–292 (1981).
20. Cook, E. R., Briffa, K. R., Shiyatov, S. G. & Mazepa, V. S. in *Methods of Dendrochronology* (eds Cook, E. R. & Kairiukstis, L. A.) 104–123 (Kluwer, Dordrecht, 1990).
21. Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A. & Funkhouser, G. The 'segment length curve' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* **5**, 229–237 (1995).
22. Leventer, A. et al. Productivity cycles of 200–300 years in the Antarctic Peninsula region: Understanding linkages among the sun, atmosphere, oceans, sea ice, and biota. *Geol. Soc. Am. Bull.* **108**, 1626–1644 (1996).
23. Stuiver, M. & Braziunas, T. F. Atmospheric ¹⁴C and century-scale solar oscillations. *Nature* **338**, 405–408 (1989).
24. Lammy, F., Hebbeln, D. & Wefer, G. Late Quaternary precessional cycles of terrigenous sediment input off the Norte Chico, Chile (27.5S) and palaeoclimatic implications. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **141**, 233–251 (1998).
25. Salinger, M. J. Palaeoclimates north and south. *Nature* **291**, 106–107 (1981).
26. Markgraf, V. et al. Evolution of late Pleistocene and Holocene climates in the circum-South Pacific land areas. *Clim. Dyn.* **6**, 193–211 (1992).
27. Boninsegna, J. A. in *Climate Since A. D. 1500* (eds Jones, P. D. & Bradley, R. S.) 446–462 (Routledge, London, 1992).
28. Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden-Day, San Francisco, 1968).
29. Marple, S. L. Jr. *Digital Spectral Analysis with Applications* (Prentice Hall, Englewood Cliffs, New Jersey, 1987).
30. Burg, J. P. in *Modern Spectrum Analysis* (ed. Childers, D. G.) 42–48 (IEEE Press, New York, 1978).
31. Lara, A. et al. in *Dendrochronology in Latin America* (ed. Roig, F. A.) 217–244 (EDIUNC, Mendoza, 2000).

Acknowledgements

We thank J. Betancourt, D. Stahle, V. Markgraf and J. Pilcher for comments on the manuscript. AMS dates were provided in part through funding by NSF (Earth System History) and the Inter-American Institute (IAI).

Correspondence and requests for materials should be addressed to F.A.R. (email: froig@lab.cricyt.edu.ar).

Simulating the amplification of orbital forcing by ocean feedbacks in the last glaciation

M. Khodri, Y. Leclainche, G. Ramstein, P. Braconnot, O. Marti & E. Cortijo

LSCE, CE-Saclay, 91191, Gif sur Yvette, France

According to Milankovitch theory, the lower summer insolation at high latitudes about 115,000 years ago allowed winter snow to persist throughout summer, leading to ice-sheet build-up and glaciation¹. But attempts to simulate the last glaciation using global atmospheric models have failed to produce this outcome when forced by insolation changes only^{2–5}. These results point towards the importance of feedback effects—for example, through changes in vegetation or the ocean circulation—for the amplification of solar forcing^{6–9}. Here we present a fully coupled ocean–atmosphere model of the last glaciation that produces a build-up of perennial snow cover at known locations of ice sheets during this period. We show that ocean feedbacks lead to a cooling of the high northern latitudes, along with an increase in atmospheric moisture transport from the Equator to the poles. These changes agree with available geological data^{10–15} and, together, they lead to an increased delivery of snow to high northern latitudes. The mechanism we present explains the onset of glaciation—which would be amplified by changes in vegetation—in response to weak orbital forcing.

Previous simulations using atmosphere general circulation models have shown that the atmosphere alone is unable to produce perennial snow, which is required to start a glaciation^{2–7,16}; several studies^{3,16} related this failure to model resolution and snow parametrization. However, a major drawback of these simulations was that the ocean surface temperature and vegetation cover were prescribed to their modern values^{2–9}. A previous study showed that biosphere–atmosphere interactions help to create favourable conditions for ice-sheet growth⁹, but did not succeed in simulating perennial snow cover. For the period of the onset of the last glaciation, studies of planktonic and benthic foraminifera from the North Atlantic Ocean led to the conclusion that the meridional gradient of sea surface temperature (SST) was enhanced before glacial inception—with very cold SST at high latitudes¹⁰ (a cooling of around 4 °C from present-day values) while low latitudes were slightly warmer^{10,12} than at present. These studies^{10,14,17–19} assumed that the changes in SST pattern were probably due to a shift of deep-water formation from the Norwegian Sea to the North Atlantic. Therefore changes in ocean circulation must have had a large effect on the high-latitude heat budget. These hypotheses remain untested by coupled ocean–atmosphere general circulation models. We tested this possible mechanism for glacial inception using the IPSL-CM2²⁰ ocean–atmosphere coupled model (see Methods section).

We performed a 100-year-long sensitivity experiment, called here the '115 kyr BP' run, which only differs from the 120-year control experiment in the Earth's orbital parameters, which are set for 115,000 years before present (BP)²¹ (Table 1). The concentration of

atmospheric CO₂ is kept at its present level, as found in the Vostok ice-core record²². Both simulations are performed without flux corrections and are stable. The changes in the orbital parameters (decreased axial tilt and increased eccentricity) reduce the amplitude of the seasonal cycle of incoming solar radiation at the top of the atmosphere, and produces a 6% decrease of summer insolation over the Northern Hemisphere, compared to the present (Fig. 1a).

These changes in insolation lead to a mean winter surface warming of 3 °C and a summer cooling of 2 °C over the Northern Hemisphere. As shown for the mid-Holocene²⁰ (6 kyr BP), the seasonal surface air temperature over land mimics the insolation

forcing with no lag in time (Fig. 1b), whereas the response of the ocean surface temperature is delayed by about 2–3 months (Fig. 1c). This time lag enhances the Equator-to-pole thermal gradient in summer, particularly over the North Atlantic where subtropical latitudes are warmer and high latitudes colder (Fig. 2a, b). The insolation changes and the colder surface ocean temperatures in the northern high latitudes also prevent the melting of sea ice in

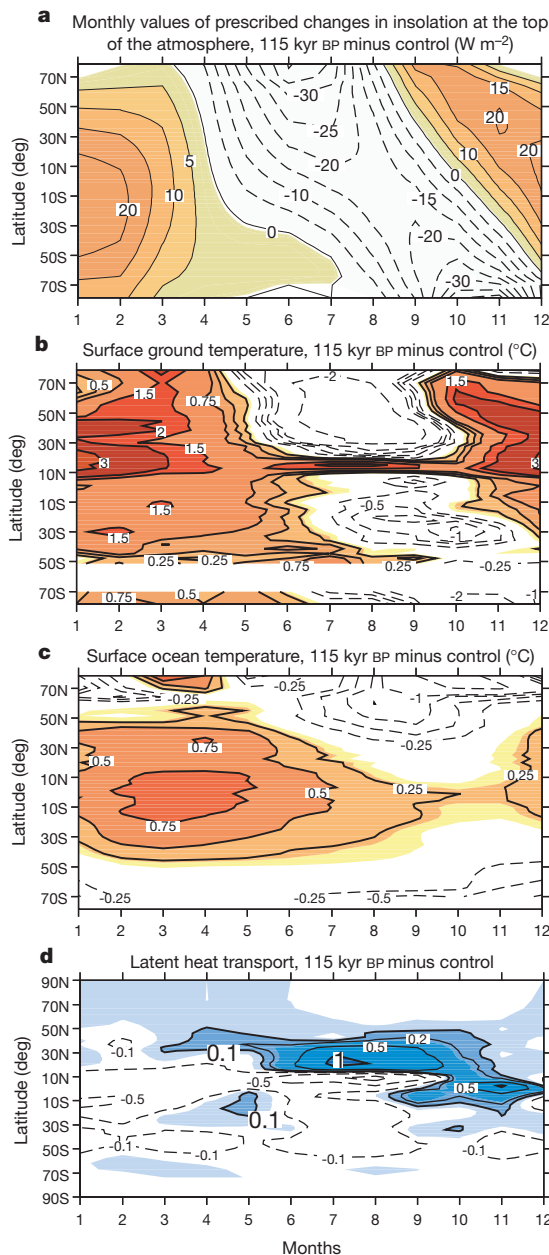


Figure 1 Seasonal variations of the response of surface temperature and the hydrological cycle to changing insolation forcing at the top of the atmosphere. **a**, **b**, Zonal-mean difference between 115 kyr BP and control simulations: **a**, for incoming solar radiation at the top of the atmosphere; **b**, for ground surface temperature; **c**, for sea surface temperature; and **d**, for northward latent heat transport over the Atlantic Ocean. Contour intervals are every 5.0 W m⁻² in **a**, every 0.25 °C in **b** and **c**, and every 0.1 PW (10¹⁵ W) in **d**, with negative values as dotted lines in all panels. Month 1 is January.

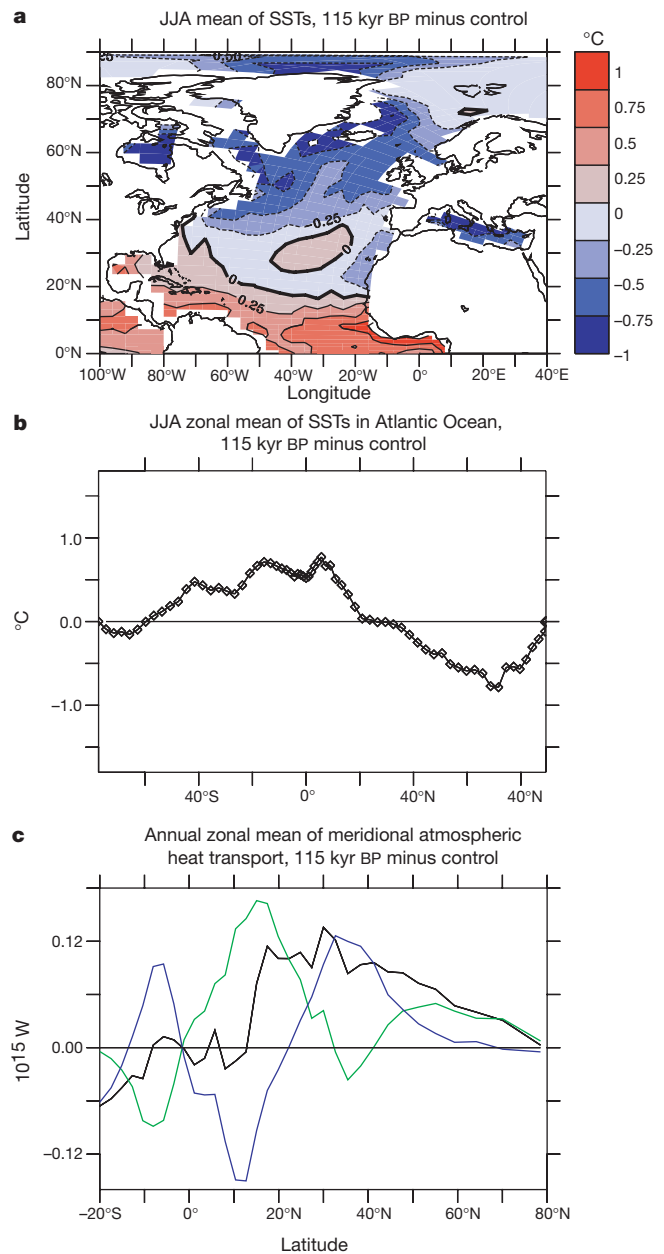


Figure 2 Changes in sea surface temperature and heat transport in response to seasonal perturbation in solar forcing. **a**, Difference between June–August means of sea surface temperature (SST) over the North Atlantic in the 115 kyr BP and control simulations. Contour intervals are every 0.25 °C, with negative values as dashed lines. **b**, Difference between June–August zonal means of Atlantic sea surface temperature in the 115 kyr BP and control simulations. The enhanced Atlantic Ocean surface temperature gradient in the Northern Hemisphere summer in the 115 kyr BP experiment compared to the control run is robust, considering the low amplitude of the inter-decadal variability of SST in the control experiment. **c**, Zonal mean difference between the 115 kyr BP and control simulations of the Equator-to-pole atmospheric heat transport in PW, for total (black line), dry static (green line) and latent (blue line). We note that there is a 6% increase of northward latent-heat transport in comparison with the control run.

summer, and lead to an increased mean sea-ice volume in the Northern Hemisphere (from $21 \times 10^3 \text{ km}^3$ in the control experiment, to $30 \times 10^3 \text{ km}^3$ for the 115 kyr BP run), which constitutes a further feedback for high-latitude cooling. The amplified summer gradient over the North Atlantic Ocean is in good agreement with SSTs inferred from marine data^{10,12,14}. The annual mean temperature shows the continents in the Northern Hemisphere to be 1°C warmer, except for northeastern Canada, Greenland, western Europe and northern Siberia, which are about 0.5°C cooler. This annual mean surface temperature pattern, which results from a

seasonal perturbation with marked latitudinal structure, is reasonable when compared to data for the end of the last interglacial^{23,24}.

The thermohaline circulation of the ocean is also affected by the orbital forcing. In the control simulation, deep-water convection occurs during winter in the Norwegian–Greenland seas and to the south and west of Iceland, with a maximum reached in March²⁵. The deep sinking of surface water at around 60°N is associated with the northward advection of relatively warm and salty waters by the North Atlantic drift to the Norwegian–Greenland–Iceland seas (GIN seas). In the 115 kyr BP experiment, the convection is less active in the GIN seas as shown by the shallower mixed-layer depth there (Fig. 3c). This deep mixing is however slightly shifted southward in the North Atlantic. Deep mixing also disappears in March, leading to a shortened season of deep-water formation. The North Atlantic meridional overturning at 2,000 m depth is therefore weakened by about 2 Sv, and the North Atlantic Deep Water (NADW) is shallower (Fig. 3a, b), confirming that the thermohaline circulation is very sensitive to external forcings²⁶. However, the maximum of the Atlantic overturning function is the same as the value in the control simulation (about 14 Sv), showing that the thermohaline circulation is still active—in agreement with data¹⁵.

The spreading of deep convection in the North Atlantic, with formation of much shallower NADW—which has been much discussed^{10,14}—appears to be induced, in our simulation, by two mechanisms. First, the summer increase of the Equator-to-pole surface temperature gradient (Fig. 2a, b) acts to enhance the annual northward transport of moisture by the atmosphere (Figs 1d, 2c). Precipitation is therefore enhanced over land, leading to a 14% annual mean increase of the Siberian river runoff into the Arctic Ocean, which in turn reduces the salinity (-0.2‰) of waters advected to the sinking sites in winter. This southward extension of the cold and less-salty Greenland current from the northern seas in winter, at the expense of the northward extension of the warm water, causes the reduction of surface salinity and temperature south of the Denmark Strait (near 60°N) despite the warmer North Atlantic induced by primary solar forcing. Second, the warmer winter continent surface together with this local winter cooling of surface waters attenuates the atmospheric stationary waves in January and February. The consequence is a decreased Icelandic low and a reduction of the pressure gradient between Iceland and Spitzberg over the Atlantic Ocean. The associated weaker cyclonic activity reduces the northeast extension of the warmer and saltier water from the North Atlantic into the Norwegian Sea. Consequently this favours the southward shift of the polar front and the reduction of water exchange between the North Atlantic and the Arctic Ocean, and therefore favours increased deep mixing in the North Atlantic at the expense of the GIN seas.

The ocean therefore induces strong positive feedbacks to the primary insolation forcing. It amplifies the initial response of surface temperature to the meridional gradient of insolation, both in latitude and time (Fig. 1a, c). This increases the annual northward atmospheric heat transport by about 4% (Fig. 2c). The maximum increase of atmospheric heat transport to the north is primarily due to latent heat transport (Fig. 2c), enhancing water transport over high latitudes of the Northern Hemisphere. The response of the

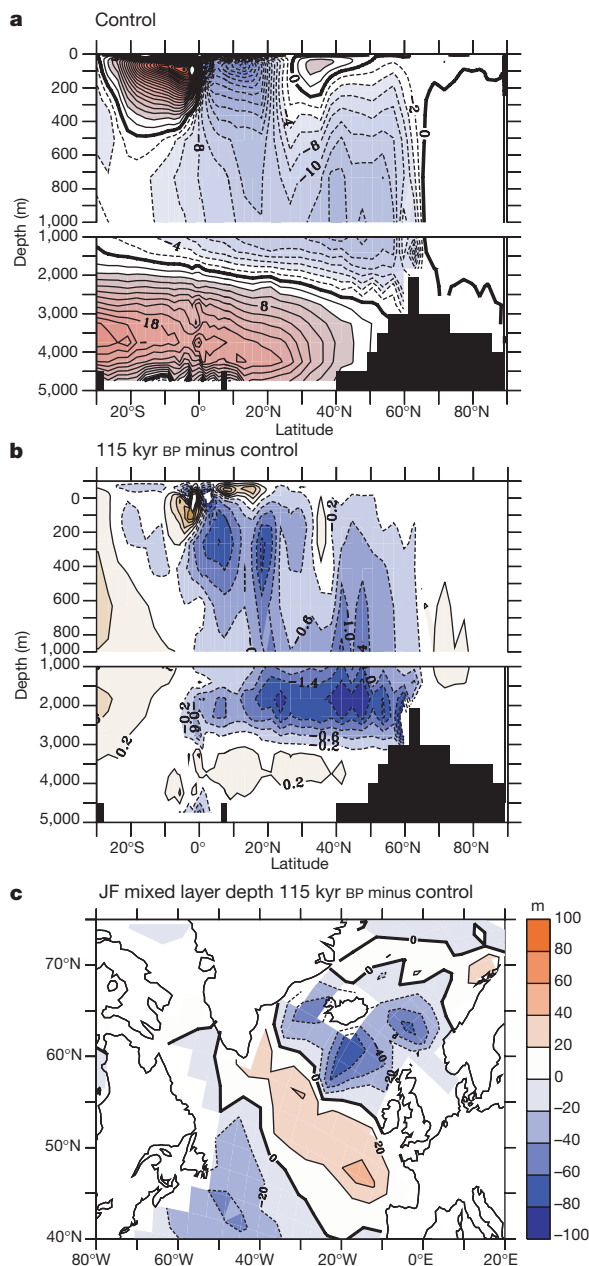


Figure 3 Response of thermohaline circulation and deep convection to changing insolation forcing. **a, b**, Latitude–depth distribution of the annual mean Atlantic stream function in sverdrups (Sv) for **a**, the control experiment and **b**, the difference between 115 kyr BP and control. **c**, Difference between January–February means of the mixed layer depth in the North Atlantic for the 115 kyr BP and control simulations. Contour intervals are every 20 m, with negative values shown as dashed lines. We note the deepening of the mixed layer depth in the North Atlantic Ocean (at 50°N) where active deep-water flow takes over from the Norwegian/Greenland sinking site (at 60°N).

Table 1 Orbital parameters of the Earth used in the coupled simulation

	Control	115 kyr BP
Eccentricity	0.016724	0.041421
Obliquity (axial tilt, $^\circ$)	23.446	22.404
Perihelion -180° ($^\circ$)	102.04	110.88
Integration time (yr)	120	100

These parameters were used for calculation of incoming solar radiation at the top of the atmosphere for the control and for the 115 kyr BP experiment. Time integration: for both simulations the oceanic drift in terms of global mean SST is about 0.25°C per century. The similarity of the drift for those simulations allows the comparison of the two resulting climates. Our analyses are based on a 70-year running mean of annual cycles.

Atlantic thermohaline circulation, in contrast to that of the atmosphere, is not directly to the amplified meridional solar forcing. By analogy with the circulation of the modern Atlantic, which acts to reduce the Equator-to-pole temperature gradient, we would expect an increased poleward ocean heat flux in the 115 kyr BP configuration. However, the equilibrium response of the NADW to changes in the winds and the hydrological cycle is such that the oceanic heat transport from low to high latitudes decreases by about 6%.

The enhanced high-latitude cooling of the Northern Hemisphere, together with the increased atmospheric moisture transport, provide optimal conditions for delivering snow over the Northern Hemisphere. This ocean-atmosphere adjustment in our simulation is the main feedback that occurs under the insolation conditions of the last glacial inception. Snow becomes perennial over the Canadian archipelagos. In these regions, snow accumulates through time in the 115 kyr BP run for two reasons. First, because of the increased summer snowfall (by about 63.5%), and second, because of the persistence of snow in summer (due to the cooling and the slight northward shift of the circulation of the atmosphere; Fig. 4). There is also a tendency for snow accumulation in northern Norway and in northeastern Europe, which are thought to be the nucleation sites for the Fennoscandian ice sheet^{27,28}. Moreover, perennial snow cover simulated over the Canadian Arctic islands is consistent with data²⁹, according to which initial snow accumulation over Fennoscandia began after the ice sheets had built-up over North America.

The role of ocean dynamics—through the rearrangement of North Atlantic deep convection—in triggering the last glacial inception is commonly assumed when interpreting palaeoclimate data^{10,14,17–19}. The main strength of our coupled model simulation is that it illustrates the role of the ocean in modifying the seasonal insolation forcing factor, both in time and amplitude. This produces significant changes in surface temperatures and deep-water production, leading to perennial snow over northeastern Canada, in good agreement with data. We have shown that in simulations, before accumulating snow and then triggering the albedo feedback, it is necessary to account for complex feedback mechanisms involving atmospheric winds, the hydrological cycle and ocean dynamics, which allow an increase of snow delivery and persistence in summer over high latitudes. We did not consider biosphere-atmosphere interactions; but such interactions would lead to further cooling by enhancing the surface albedo of high latitudes⁹, thus favouring the onset of glaciation. For this reason, biosphere-atmosphere interactions need to be considered in future work. Our results confirm the considerable effects that modifications of ocean dynamics can have on the climate of the Northern Hemisphere middle and high latitudes; this may help to shed light on the potential role of ocean dynamics in future climate changes. □

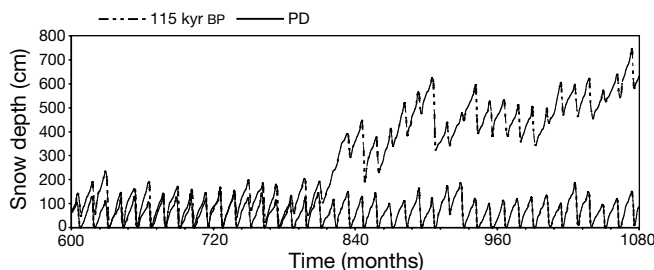


Figure 4 Monthly time series of snow depth in centimetres. Time starts from January of year 50 for the control experiment (solid line) and for the 115 kyr BP experiment (dashed line) at a glaciation-sensitive region in the Laurentide area (70°N; 80°W). We see that seasonality is important. Snow depth is very stable all along the control run, with no snow in summer, whereas the 115 kyr BP run shows an increase of snow depth associated with perennial snow cover beginning at year 80 and lasting until the end of the run.

Methods

The IPSL-CM2 ocean-atmosphere coupled model is an improved version of the IPSL-CM1 ocean-atmosphere coupled model²⁰. The atmospheric model has a resolution of about 400 × 400 km at 50° latitude. The atmospheric component includes an improved version of the SECHIBA land-surface scheme²⁰. The ocean component horizontal mesh is orthogonal and curvilinear on the sphere. The northern point of the convergence has been shifted onto Asia, in order to overcome the singularity of Northern Hemisphere high latitudes. The ocean model spatial resolution over high latitude reaches a maximum size of 4° by 3°. This version of the model includes a complex sea-ice model, with three different ice types having specific thermodynamic behaviour²⁵. A variable fraction of leads in the sea-ice cover is simulated. Dynamical processes are however not represented. The Northern Hemisphere sea-ice cover and seasonal cycle is realistic. Maximum ($13.7 \times 10^6 \text{ km}^2$) and minimum ($7.5 \times 10^6 \text{ km}^2$) ice extent are in the range of satellite estimates ($13.5 \times 10^6 \text{ km}^2$ and $6 \times 10^6 \text{ km}^2$ respectively²⁵). In the present-day climate simulation, this complex thermodynamic sea-ice model has a large effect on the characteristics of water masses in the Arctic, in the GIN seas, and over deep convection. This model controls in part the seasonality of surface water salinity, as it simulates brine rejection during ice formation and release of fresh water during ice melting. The higher salinity in winter together with the cooling allows deep convection over the GIN seas, and a deep-water flow through the Denmark Strait of about 6.1 Sv—which is in good agreement with the estimates of about 6 Sv (ref. 30). In turn, this convection strengthens surface advection, and the annual mean maximum of the overturning function in the North Atlantic, which is about 14.6 Sv. Accumulation of snow, its transformation into ice due to snow flooding, the evolution of snow surface albedo with snow age, and specific albedos for frozen and melted ice, are all represented in the model. The two models exchange ocean and sea-ice surface temperatures, sea-ice cover, sea-ice albedo, momentum, heat and fresh water fluxes once a day²⁰. The 115 kyr BP experiment is a 100-year simulation, starting from year 40 of the control run by simply switching the Earth's orbital parameters to those valid for 115 kyr BP. For both simulations, the oceanic drift in terms of SST is only about 0.22 °C per century, despite the no-flux correction.

Precipitation falls as snow in the model, if the air temperature above the surface is less than 0 °C. Snow cover and snow age affect the surface albedo over land and the heat capacity of the surface. Albedos due to snow cover also vary with surface type (with or without vegetation). Sublimation of snow is calculated as part of the surface evaporative flux, and snowmelt contributes to soil moisture. At the surface of snow-covered grid points, the net heat gain is used to heat snow until it reaches the freezing point (0 °C) and the excess heat gain is used to melt snow. Our modern climate experiment with this snow parametrization shows snow cover, repartition and seasonal variation in the Northern Hemisphere in good agreement with climatology, which is a necessary precondition for glacial inception studies.

Received 10 July 2000; accepted 22 January 2001.

- Milankovitch, M. K. Kanon der erdbestrahlung und seine anwendung auf daseiszeitenproblem. *Serb. Acad. Beogr. Spec. Publ.* **132**, (1941) (in Yugoslavian); Canon of Insolation and the Ice Age Problem (Israel Program for Scientific Translation, Jerusalem, 1969) (English transl.).
- Royer, J. F., Deque, M. & Pestiaux, P. Orbital forcing of the inception of the Laurentide ice sheet? *Nature* **304**, 43–46 (1983).
- Rind, D., Peteet, D. & Kukla, G. Can Milankovitch orbital variations initiate the growth of ice sheets in a general circulation model? *J. Geophys. Res.* **94**, 12851–12871 (1989).
- Oglesby, R. J. Sensitivity of glaciation to initial snow cover, CO₂, snow albedo, and oceanic roughness in the NCAR GCM. *Clim. Dyn.* **4**, 219–235 (1990).
- Mitchell, J. F. B. Modelling of paleoclimates: examples from the recent past. *Phil. Trans. R. Soc. Lond. B* **341**, 267–275 (1993).
- Dong, B. & Valdes, P. J. Sensitivity studies of northern hemisphere glaciation using an atmospheric general circulation model. *J. Clim.* **8**, 2471–2495 (1995).
- Phillips, P. J. & Held, M. The response to orbital perturbations in a atmospheric model coupled to a slab ocean. *J. Clim.* **7**, 767–782 (1994).
- Gallimore, R. G. & Kutzbach, J. E. Role of orbitally induced changes in tundra area in the onset of glaciation. *Nature* **381**, 503–505 (1996).
- DeNoblet, N. *et al.* Possible role of atmosphere-biosphere interactions in triggering the last glaciation. *Geophys. Res. Lett.* **23**, 3191–3194 (1996).
- Cortijo, E. *et al.* Changes in meridional temperature and salinity gradients in the North Atlantic Ocean (30°–72°N) during the last interglacial period. *Paleoceanography* **14**, 23–33 (1999).
- Duplessy, J. C. & Shackleton, N. J. Response of global deep-water circulation to Earth's climatic change 135,000–107,000 years ago. *Nature* **316**, 500–507 (1985).
- Ruddiman, W. F. & McIntyre, A. Warmth of the subpolar North Atlantic ocean during Northern Hemisphere ice-sheet growth. *Science* **204**, 173–175 (1979).
- Adkins, J. F., Boyle, E. A., Keigwin, L. & Cortijo, E. Variability of the North Atlantic thermohaline circulation during the last interglacial period. *Nature* **390**, 154–156 (1997).
- Cortijo, E. *et al.* Eemian cooling in the Norwegian Sea and North Atlantic ocean preceding continental ice-sheet growth. *Nature* **372**, 446–449 (1994).
- Duplessy, J. C. & Shackleton, N. J. Deepwater source variations during the last climatic cycle and their impact on deep water circulation. *Paleoceanography* **3**, 343–360 (1988).
- Marshall, S. J. & Clarke, G. K. C. Ice sheet inception: Subgrid hypsometric parameterization of mass balance in an ice sheet model. *Clim. Dyn.* **15**, 533–550 (1999).
- Imbrie, J. *et al.* On the structure and origin of major glaciation cycles, 1. Linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992).
- Imbrie, J. *et al.* On the structure and origin of major glaciation cycles, 2. The 100,000-year cycle. *Paleoceanography* **8**, 669–735 (1993).
- Balbon, E. *Variabilité Climatique et Circulation Thermohaline dans l'océan Atlantique Nord et en Mer de Norvège au Cours du Dernier Quaternaire Supérieur*. Thesis, Orsay Univ. (2000).
- Braconnot, P., Marti, O., Joussaume, S. & Leclainche, Y. Ocean feedback in response to 6kyr BP insolation. *J. Clim.* **13**, 1537–1553 (2000).

21. Berger, A. Long term variation of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
22. Raynaud, D. *et al.* The ice record of greenhouse gases. *Science* **259**, 926–934 (1993).
23. LIGA members. The last interglacial in high latitudes of the northern hemisphere: Terrestrial and marine evidence. *Quat. Int.* **10–12**, 9–28 (1991).
24. Sánchez Goni, M. F., Eynaud, F., Turon, J. L. & Shackleton, N. J. High resolution palynological record off the Iberian margin: direct land-sea correlation for the last interglacial complex. *Earth Planet. Sci. Lett.* **171**, 123–137 (1999).
25. Leclainche, Y. *et al.* Diagnostic versus thermodynamic sea ice models in global coupled ocean-atmosphere simulations. *Clim. Dyn.* (submitted).
26. Rahmstorf, S. Bifurcation of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
27. Ives, J. D., Andrews, J. T. & Barry, R. G. Growth and decay of the Laurentide ice sheet and comparison with Fenno-Scandinavia. *Naturwissenschaften* **2**, 118–125 (1975).
28. Andrews, J. T. & Mahafy, M. A. W. Growth rates of the Laurentide ice sheet and sea level lowering (with emphasis on the 115,000 B.P. sea level low). *Quat. Res.* **6**, 167–183 (1976).
29. Mangerud, J. in *Klimageschichtliche Probleme der letzten 130,000 Jahre* (ed. Frenzel, B.) 307–330 (Fischer, Stuttgart, 1991).
30. Schmitz, W. J. & McCartney, M. S. On the North Atlantic circulation. *Rev. Geophys.* **31**, 29–49 (1993).

Acknowledgements

We thank the LODYC for providing the ocean model and the sea-ice model, the LMD for the atmosphere model, and the CERFACS for the coupler, OASIS. Computer time was provided by the Commissariat à l'Energie Atomique. We thank S. Harrison, J. C. Duplessy and D. Paillard for their comments, which helped to improve the manuscript.

Correspondence and requests for materials should be addressed to M.K.
(e-mail: khodri@lscce.saclay.cea.fr).

Late Jurassic salamanders from northern China

Ke-Qin Gao* & Neil H. Shubin†

* American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

† Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, Illinois 60637, USA

With ten extant families, salamanders (urodeles) are one of the three major groups of modern amphibians (lissamphibians)^{1–6}. Extant salamanders are often used as a model system to assess fundamental issues of developmental, morphological and biogeographical evolution^{6–11}. Unfortunately, our understanding of these issues has been hampered by the paucity of fossil evidence available to assess the early history of the group^{5,6,12}. Here we report the discovery of an extraordinary sample of salamander fossils, some with rare soft-tissue impressions, from the Upper Jurassic of China^{13–16}. With over 500 articulated specimens, this assemblage documents the morphological diversity of early urodeles and includes larvae and adults of both neotenic and metamorphosed taxa. Phylogenetic analysis confirms that these salamanders are primitive, and reveals that all basal salamander clades have Asian distributions. This is compelling evidence for an Asian origin of Recent salamanders, as well as for an extensive and early radiation of several major lineages. These discoveries show that the evolution of salamanders has involved phylogenetic and ecological diversification around a body plan that has remained fundamentally stable for over 150 million years.

Amphibia Linnaeus, 1758
Lissamphibia Haeckel, 1866
Caudata Scopoli, 1777
Urodela Dumeril, 1806
Family incertae sedis

Sinerpeton fengshanensis gen. et sp. nov.

Holotype. GMV 1606, articulated skeleton including cranium and postcranium.

Etymology. *Sino* + *herpeton* (Greek, meaning creeping animal from China); fengshan (the type locality).

Type locality and horizon. Fengshan, Hebei Province, China; Late Jurassic fossil beds that overlay the Zhangjiakou Formation (Tithonian)^{13–16}.

Referred material. GMV 1607–1610, and numerous unnumbered topotypic specimens.

Diagnosis. *S. fengshanensis* is a neotenic form that retains primitive urodele characters including: small dorsal processes on paired premaxillae; paired nasals articulated in midline; the presence of separate angular and coronoid; and a basale commune representing fused distal carpals 1+2 in the manus. It shares derived characters such as uncapitate ribs with cryptobranchoids, but differs from this group in having the following combination of characters: presacral vertebrae with laterally expanded zygapophyses; a single centrale; a greatly expanded metacarpal II in the manus; and phalangeal formulae of 1-2-3-2 in manus and 1-2-3-4-2 in pes.

The Fengshan assemblages include articulated specimens of both neotenic (*S. fengshanensis*; Figs 1, 2) and fully metamorphic salamanders (*Laccotriton subsolanus*; Fig. 3)¹⁷. The holotype of the new taxon *Sinerpeton* retains an ossified mesopodium and paired ossified ceratobranchials; the combination of these features suggests that this specimen represents a mature individual with external gills. The assessment of *L. subsolanus* as a metamorphic form is based on its high degree of ossification, particularly in the mesopodium, and apparent lack of external gills—ossified ceratobranchials are absent in adults (Fig. 3). Larvae are also present in the Fengshan sample. These specimens, some of which preserve impressions of the soft tissues of the external gills, have weakly ossified skeletons, are smaller than the metamorphosed individuals of *Laccotriton*, and retain the distinctive skull proportions seen in many larval salamanders (Fig. 3).

The 500-plus articulated salamander skeletons from the Fengshan locality were recovered from a small area of no more than 10 m². The high concentration of fully articulated specimens in tuffaceous shales indicates catastrophic mass mortality during a pyroclastic eruption. The age of the fossil beds is probably mid-Tithonian or slightly younger, on the basis of radiometric dating and biostratigraphy. The underlying Zhangjiakou Formation yields an ⁴⁰Ar/³⁹Ar

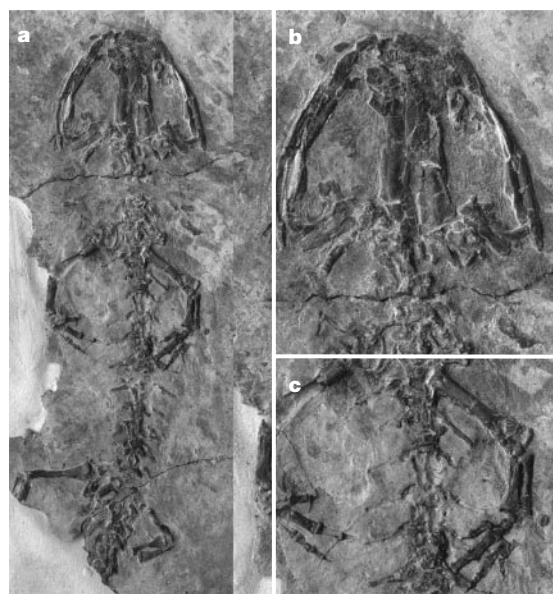


Figure 1 Holotype (GMV 1606) of *S. fengshanensis*, gen. et sp. nov. **a**, Dorsal view of articulated skeleton including the cranium and postcranium. **b**, Enlarged skull. **c**, Enlarged pectoral girdle and forelimbs.

date of 151 million years BP (ref. 13). Biostratigraphic comparisons of the insect assemblage suggest a Tithonian age¹⁴, and analysis of conchostracans^{15,16} suggests an age of Kimmeridgian. Therefore, the salamander fossils from Fengshan represent the earliest true urodeles that are known from fully articulated skeletons. In addition, some of these fossils have preserved rare soft-tissue impressions (Fig. 3a, b). Consequently, these early salamanders reveal pivotal anatomical details that are not available from other known fossils, and have fundamental evolutionary and biogeographic implications.

Sinerpeton and *Laccotriton* have cranial morphologies that are similar to those of extant hynobiids in having a lacrimal in the skull and a separate angular in the mandible. The retention of a well-defined coronoid along with other characters, however, indicates a more basal position of the two Late Jurassic taxa. In limb structure, an important trend in salamanders is the reduction in the number of mesopodial elements with derived limb patterns almost always involving the loss or fusion of bones^{9,10}. This trend was underway by the Late Jurassic: both the neotenic and metamorphosed salamanders from Fengshan possess fused distal carpals 1+2 in the manus, and distal tarsals 1+2 in the pes. An Early Cretaceous salamander from Liaoning also shows a similar pattern¹⁸, although another salamander of the same age has unossified mesopodials¹⁹. Indeed, this character is one of the most diagnostic features of Recent salamanders and has remained remarkably stable for the past 150 million years in the evolution of urodeles.

Inclusion of the Fengshan material in analyses of salamander phylogeny provides new insights into the relationships and biogeography of the main salamander clades. Most recent phylogenetic hypotheses recognize the neotenic family Sirenidae as representing the most basal extant salamanders^{1,3,20–23}. Our analyses of morphological data independently and in combination with molecular data²² suggest that sirenids are not basal but are most closely related

to another neotenic clade, the proteids (Fig. 4). To move sirenids to a basal position involves the addition of 26 and 16 steps to the morphological and combined trees, respectively. The analyses also reveal that Asian taxa lie at the base of caudate phylogeny: both *Sinerpeton* and *Laccotriton* are placed at the base of Urodela clade together with Cryptobranchioidea. As hynobiids are almost exclusively Asian in distribution, and cryptobranchids have a combined Asian/North American distribution, the most parsimonious hypothesis is that the basal salamanders initially radiated in Asia, and later spread to North America and Europe. This new hypothesis is further supported by the presence of a basal caudate, *Karaurus*, in Late Jurassic deposits of Kazakhstan²⁴, and a younger record of salamanders from the Early Cretaceous of both Europe and Asia^{12,18,19}. All other Jurassic records of salamanders consist of isolated elements and are either taxonomically equivocal or not amenable to phylogenetic analysis^{25,26}.

Analyses of salamander phylogeny are plagued by both morphological conservatism and homoplasy of major characters. The study of the new Fengshan material suggests that these phylogenetic difficulties reflect the biological realities of the salamander cladogenesis. Phylogenetic analysis of fossil evidence strongly supports the hypothesis that major extant salamander families have been

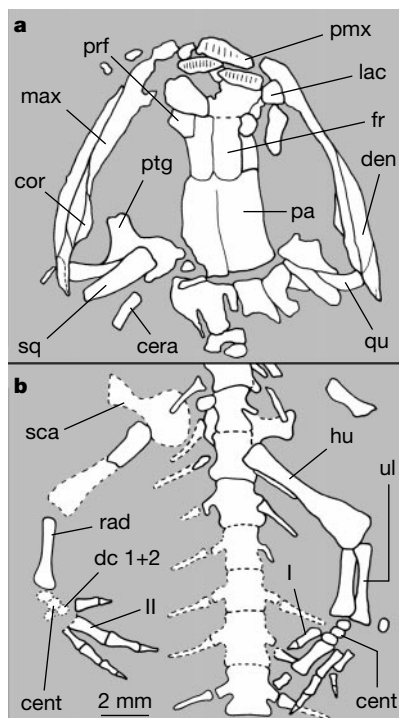


Figure 2 Line drawing of the holotype of *S. fengshanensis*, gen. et sp. nov. **a**, Dorsal view of skull. **b**, Pectoral girdle and forelimbs. cent, centrale; cera, ceratobranchial; cor, coronoid; dc, distal carpal; den, dentary; fr, frontal; hu, humerus; lac, lacrimal; max, maxilla; pa, parietal; pmx, premaxilla; prf, prefrontal; ptg, pterygoid; qu, quadrate; rad, radius; sca, scapulocoracoid; sq, squamosal; ul, ulna.

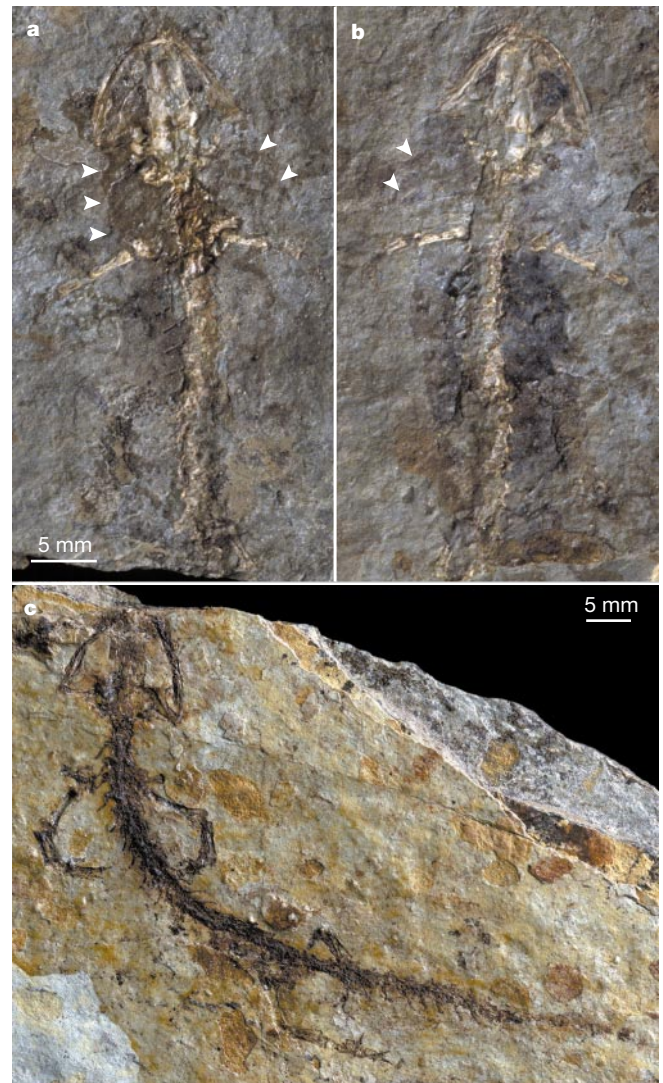


Figure 3 *L. subsolanus*. **a**, **b**, Part and counterpart of a juvenile specimen (GMV 1603) with arrows pointing to impressions of external gills. **c**, GMV 1602 (holotype), nearly complete cranial and postcranial skeleton and impressions.

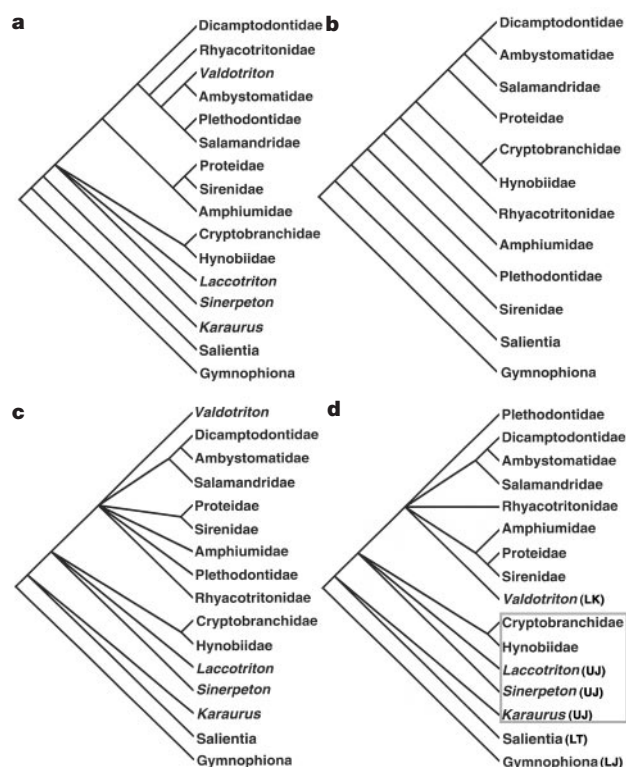


Figure 4 Phylogenetic relationships of major salamander families. **a**, Strict consensus of the nine most parsimonious trees derived from morphological data (length, 109; consistency index (CI), 0.629; retention index (RI), 0.713). **b**, Single tree (length, 258; CI, 0.578; RI, 0.579) from molecular data in ref. 22. **c**, Strict consensus of 21 trees (length, 402; CI, 0.549; RI, 0.537) derived from combined analysis of morphological and molecular data with insertion/deletion events coded as a fifth state. **d**, Strict consensus of nine trees (length, 297; CI, 0.556; RI, 0.584) from combined analysis with insertion/deletion events coded as missing data. LJ, Lower Jurassic; LK, Lower Cretaceous; LT, Lower Triassic; UJ, Upper Jurassic.

evolving separately since at least the Early Cretaceous (Fig. 4). Extant salamander families are characterized by a diversity of life-history strategies. Neoteny, in particular, has been a key component in the generation of salamander diversity; over 40 species in 9 different families show this evolutionary strategy²⁰. The Fengshan salamanders reveal that this major mode of urodele evolution was already established by the Late Jurassic. The early and independent occurrences of neoteny in Mesozoic and Recent salamander groups make parallel evolution of numerous morphological features a long-standing factor of salamander history. □

Methods

Phylogenetic analyses

We used PAUP version 3.1.1 to analyse the data sets (see Supplementary Information) and Branch-and-Bound search algorithm to identify most parsimonious trees. We treated all characters unordered and equally weighted, and rooted the trees using *Karaurus*, *Salientia* and *Gymnophiona* as successive outgroups. We did not include the Early Cretaceous *Liaoxitriton* and *Jeholotriton* in this analysis because the preliminary description of these taxa did not provide sufficient character information to evaluate their placement in a phylogenetic framework.

Description of *Sinerpeton*

The holotype is a well-preserved skeleton with a snout-vent length of 47 mm. The maxilla contacts both a small lacrimal and a reduced prefrontal dorsally. The frontals are short, having an essentially transverse anterior suture with the nasals and a roughly W-shaped posterior suture with the parietals. The parietals lack an anterolateral extension, and form the medial border of the orbit with the frontals. The squamosal in dorsal view forms a transverse bar that contacts the parietal medially. The pterygoid is primitively tri-radiate and boomerang-shaped. The mandible has five separate elements, including the dentary,

prearticular, angular, coronoid and articular.

The vertebral column consists of 16–17 presacral, a sacral and more than 20 caudals. A pair of spinal nerve foramina penetrates the atlas laterally, but these are not present in other vertebrates. The presacral vertebrae have laterally expanded zygapophyses, differing from the unexpanded condition in cryptobranchoids. Ribs are unicipitate with an expanded head. Up to four anterior caudals are also rib-bearing. In the manus, the distal carpals are well ossified, and possess the standard urodele fusion of distal carpals 1+2. The phalangeal formulae are 1-2-3-2 in the manus, and 1-2-3-4-2 in the pes.

Redescription of *Laccotriton*

Laccotriton is a metamorphic form, having a well-ossified mesopodium but lacking ossification of the ceratobranchials. It shares with cryptobranchoids derived characters such as unicipitate ribs with a broadened base, but differs from this group in having more than three pairs of caudal ribs, and the absence of spinal nerve foramina on presacral and caudal vertebrae. It differs from neotenic *Sinerpeton* in having two centralia in the manus, and phalangeal formulae of 2-2-3-2 in the manus and 2-2-3-4-2 in the pes.

The skull of *Laccotriton* retains a small lacrimal anterolateral to the prefrontal. The frontals are short, lacking the anterior extension seen in cryptobranchoids, sirenids and amphiumids. The squamosal is a transverse bar that connects the parietal to the quadrate. The vomers are paired and triangular, with vomerine teeth forming a short arcade that runs mediolaterally. The pterygoid is triradiate, lacking the anteromedial process seen in cryptobranchoids. As in *Sinerpeton*, five separate mandibular elements are present (dentary, prearticular, angular, coronoid and articular).

The vertebral column comprises 16 presacral, a single sacral and more than 20 caudal vertebrae. No spinal nerve foramina are identifiable in the vertebrae. All of the presacral but the atlas carry short and rod-like ribs, with a unicipitate and slightly expanded proximal head. At least five anterior caudals bear free ribs.

Received 10 October; accepted 21 December 2000.

- Milner, A. R. in *Amphibian Biology* (eds Heatwole, H. E. & Carroll, R. L.) 4, 1412–1444 (Surrey Beatty & Sons, Chipping Norton, Australia, 2000).
- Trueb, L. & Cloutier, R. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H. -P. & Trueb, L.) 223–313 (Cornell Univ. Press, Ithaca, London, 1991).
- Cannatella, D. C. & Hillis, D. M. Amphibian relationships: phylogenetic analysis of morphology and molecules. *Herpetol. Monogr.* 7, 1–7 (1993).
- Feller, A. E. & Hedges, B. Molecular evidence for the early history of living amphibians. *Mol. Phylog. Evol.* 9, 509–516 (1998).
- Estes, R. *Handbuch der Paläoherpetologie. Teil 2. Gymnophiona, Caudata* (Gustav Fischer, Stuttgart, 1981).
- Milner, A. R. in *Evolution, Time and Space: the Emergence of the Biosphere* (eds Sims, R. W., Price, J. H. & Whalley, P. E. S.) 431–468 (Academic, London, 1983).
- Wake, D. B. & Larson, A. Multidimensional analysis of an evolving lineage. *Science* 238, 42–48 (1987).
- Wilbur, H. M. & Collins, J. P. Ecological aspects of amphibian metamorphosis. *Science* 182, 1305–1314 (1973).
- Shubin, N. & Wake, D. B. Phylogeny, variation, and morphological integration. *Am. Zool.* 36, 51–60 (1996).
- Shubin, N., Wake, D. B. & Crawford, A. J. Morphological variation in the limbs of *Taricha granulosa* (Caudata: Salamandridae): evolutionary and phylogenetic implications. *Evolution* 49, 874–884 (1995).
- Wake, D. B. Homoplasy: the result of natural selection or evidence design limitations? *Am. Nat.* 138, 543–567 (1991).
- Evans, S. E. & Milner, A. R. A metamorphosed salamander from the Early Cretaceous of Las Hoyas, Spain. *Phil. Trans. R. Soc. Lond. B* 351, 627–646 (1996).
- Luo, X. & Li, P. A study on the boundary age between Jurassic and Cretaceous. *Acta Geol. Sinica* 18, 242–247 (1997).
- Ren, D., Yin, J. & Huang, B. Preliminary research on Late Mesozoic insect communities and ecostratigraphy of Fengning, Hebei Province. *Geol. Sci. Tech. Inform.* 18, 39–44 (1999).
- Chen, P. Nonmarine Jurassic strata of China. *Mus. North. Arizona Bull.* 60, 395–412 (1996).
- Wang, S. Correlation of the continental Jurassic in the north of China to the paralic Jurassic in northwest Scotland, United Kingdom. *Acta Geol. Sinica* 72, 11–21 (1998).
- Gao, K., Cheng, Z. & Xu, X. First report of Mesozoic urodeles from China. *Chin. Geol.* 248, 40–41 (1998).
- Dong, Z. & Wang, Y. A new urodele (*Liaoxitriton zhongjiani* gen. et sp. nov.) from the Early Cretaceous of western Liaoning Province, China. *Vert. Palasiat.* 36, 159–172 (1998).
- Wang, Y. A new salamander (Amphibia: Caudata) from the Early Cretaceous Jehol Biota. *Vert. Palasiat.* 38, 100–103 (2000).
- Duellman, W. E. & Trueb, L. *Biology of Amphibians* (McGraw-Hill Book Company, New York, 1986).
- Larson, A. in *Evolutionary Biology* (eds Hecht, M. K., Wallace, B. & MacIntyre, R. J.) 25, 211–277 (Plenum, New York, 1991).
- Larson, A. & Dimmick, W. W. in *Amphibian relationships: Phylogenetic analysis of morphology and molecules.* (eds Cannatella, D. & Hillis, D.) *Herpetol. Monogr.* 7, 77–93 (1993).
- Sever, D. M. Comparative anatomy and phylogeny of the cloacae of salamanders (Amphibia: Caudata). *Herpetologica* 47, 165–193 (1991).
- Ivachnenko, M. F. Urodelans from the Triassic and Jurassic of Soviet Central Asia. *Paleontol. J.* 12, 362–368 (1978).
- Evans, S. E., Milner, A. R. & Mussett, F. The earliest known salamanders (Amphibia: Caudata): a record from the Middle Jurassic of England. *Geobios* 21, 539–552 (1988).
- Evans, S. E. & Milner, A. R. in *In the Shadow of the Dinosaurs* (eds Fraser, N. C. & Sues, H.-D.) 303–321 (Cambridge Univ. Press, Cambridge, 1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Wake and R. Carroll for discussion; K. Monoyios for editorial assistance; M. Ellison for illustrations; and A. Davidson for preparing GMV 1606. For help in field collection of specimens, we thank Z. Cheng, X. Xu, Y. Jin and Z. Chen. N.H.S. was supported by a grant from the John Simon Guggenheim Memorial Foundation. Field work was supported by the National Geographic Society and the Frick Funds of the American Museum of Natural History.

Correspondence and requests for materials should be addressed to K.-Q.G. (e-mail: kgao@amnh.org).

Caterpillar-induced nocturnal plant volatiles repel conspecific females

Consuelo M. De Moraes*, Mark C. Mescher† & James H. Tumlinson*

* USDA-ARS, CMAVE, PO Box 14565, Gainesville, Florida 32604, USA

† Department of Entomology, University of Georgia, Athens, Georgia 30602, USA

Plants respond to insect herbivory by synthesizing and releasing complex blends of volatile compounds, which provide important host-location cues for insects that are natural enemies of herbivores^{1–3}. The effects of these volatile blends on herbivore behaviour have been investigated to only a limited extent^{4,5}, in part because of the assumption that herbivore-induced volatile emissions occur mainly during the light phase of the photoperiod^{6,7}. Because many moths—whose larvae are some of the most important insect herbivores—are nocturnal, herbivore-induced plant volatiles have not hitherto been considered to be temporally available as host-location cues for ovipositing females. Here we present chemical and behavioural assays showing that tobacco plants (*Nicotiana tabacum*) release herbivore-induced volatiles during both night and day. Moreover, several volatile compounds are released exclusively at night and are highly repellent to female moths (*Heliothis virescens*). The demonstration that tobacco plants release temporally different volatile blends and that lepidopteran herbivores use induced plant signals released during the dark phase to choose sites for oviposition adds a new dimension to our understanding of the role of chemical cues in mediating tritrophic interactions.

Feeding by insect herbivores induces plants to release chemical signals that serve as important foraging cues for parasitoids and predators, and thus enhance the plants' defence^{1–3,8–10}. Synthesis and release of these chemical signals is an active physiological process triggered by substances in the oral secretion of herbivores^{11,12}. The recent discovery that plant volatiles can transmit herbivore-specific information that allows natural enemies to identify particular herbivore species demonstrated that chemically mediated plant–insect interactions are more sophisticated and complex than was previously appreciated¹³. However, the role of chemical signals in plant–herbivore interactions remains largely unexplored. Some researchers have examined the effects of constitutive plant volatiles and herbivore-induced daytime volatiles on conspecific herbivores^{4,14–17} including some lepidopterans¹⁸, but the effect of herbivore-induced plant volatiles on moths that are active at night has been neglected. The fact that several major terpene components of herbivore-induced plant volatiles have high emissions during the periods of maximal photosynthesis^{6,7} may explain why little attention has been paid to the importance of these volatiles to female moths searching for oviposition sites at night. To our knowledge, this study represents the first demonstration that plants emit herbivore-induced volatile blends that exhibit systematic temporal variation, that some volatile compounds are released exclusively at

night, and that female moths exploit these specific night-time signals to avoid oviposition on previously damaged plants.

Gas chromatographic analysis of volatiles collected in two-hour intervals continuously for seven days revealed consistent differences in the composition of volatile blends released by *H. virescens*-infested tobacco plants ($n = 6$) during the light and dark phases of the photoperiod (Fig. 1a). Visual and auditory observations confirmed that larvae fed during both the light and dark phases. Seven major compounds were consistently released during both light and dark phases, but usually in lesser amounts during the dark phase (Fig. 1a). In addition, five compounds ((*Z*)-3-hexenyl butyrate, (*Z*)-3-hexenyl isobutyrate, (*Z*)-3-hexenyl acetate, (*Z*)-3-hexenyl tiglate, and one unidentified compound) were produced only during the dark phase. Others—(*E*)-2-hexenal and three unidentified compounds—were produced in significantly larger amounts during the dark than the light period. Thus, the qualitative and quantitative composition of volatile blends emitted by tobacco plants in response to feeding by *H. virescens* larvae can differ significantly between night and day.

Repetition of our analysis using two other species of lepidopteran larvae ($n = 6$ per species), *Manduca sexta* and *Helicoverpa zea*, provided further evidence of induced volatile release from tobacco plants during the dark period (Fig. 1b). Although the volatile

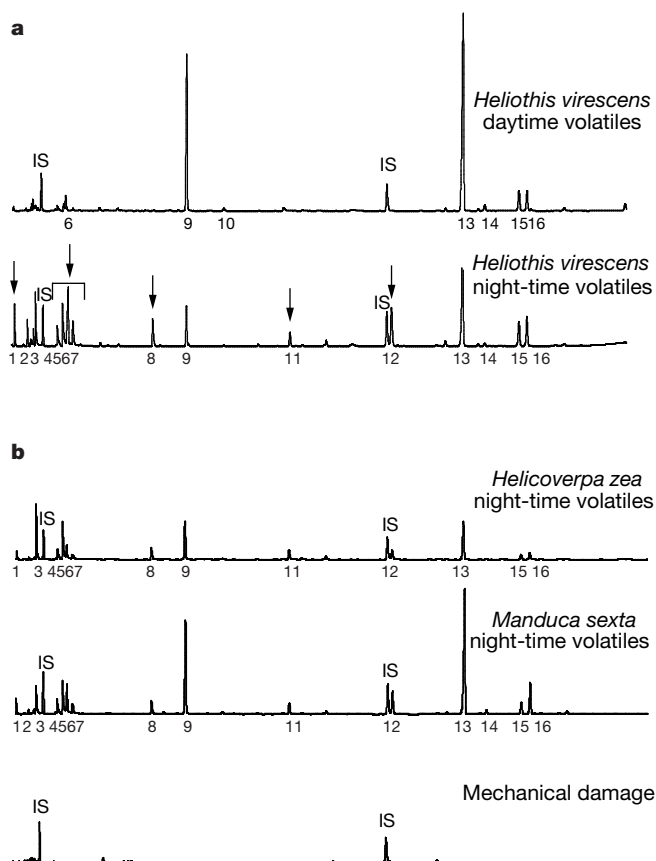


Figure 1 Gas chromatographic analysis of induced plant volatiles. **a**, Diurnal and nocturnal profiles of volatiles released from tobacco plants during a 2-h interval after 48 h of feeding by *H. virescens*. Arrows represent volatiles that are present only (or in significantly larger amounts) in the nocturnal profile. **b**, Nocturnal profiles of volatiles released from tobacco plants during a 2-h interval after 48 h of feeding by *H. virescens*, *H. zea* or *M. sexta* compared with mechanical damage. Represented are: 4, (*E*)-2-hexenal; 5, (*Z*)-3-hexen-1-ol; 8, (*Z*)-3-hexenyl acetate; 9, (*E*)- β -ocimene; 10, linalool; 11, (*Z*)-3-hexenyl butyrate; 12, (*Z*)-3-hexenyl tiglate; 13, β -caryophyllene; 14, α -humulene; 15, (*E,E*)- α -farnesene; 16, unidentified sesquiterpene; compounds 1–3, 6, 7 are unidentified compounds; IS, internal standards (*n*-octane and *n*-nonyl-acetate).

profiles revealed quantitative differences in plant response to the three herbivores, all caterpillar species induced release of the same volatile compounds from tobacco during the night (Fig. 1). To determine whether plants continue nocturnal volatile production in the absence of continuous feeding by the caterpillars, *H. virescens* were removed from the plants at 15:00 after 48 hours of feeding. In this case, plants still emitted volatiles during the dark phase although in smaller amounts.

In behavioural trials, mated *H. virescens* females spent a significantly greater proportion of time (80% of the observed hour) in an area with only undamaged plants than in an area with both damaged and undamaged plants (Fig. 2A, a). Moths exhibited a tendency to fly south when first released; however, if the damaged plants were placed to the south, the moths would often turn and fly in the opposite direction. The same preference was displayed in oviposition. Females selected only non-infested plants for oviposition (Fig. 2A, b; $F_{1,10} = 427$, $P < 0.0001$) and also avoided uninfested plants close to infested plants.

Because the preference of *H. virescens* for uninfested plants might conceivably be explained by visual cues associated with plant damage, we repeated the behavioural assays using synthetic volatile blends. Synthetic blends of the major volatile compounds that were produced in significant amounts and emitted by the plants during the dark period were formulated on rubber septa^{19,20} to release volatiles in approximately the same proportions and amounts as released by herbivore-damaged plants. In behavioural assays, *H. virescens* demonstrated a significant preference ($F_{1,10} = 134$,

$P < 0.0001$) for untreated tobacco plants over those with septa releasing the synthetic blends (Fig. 2B, b). In control trials, the response to plants with septa containing only solvent (hexane) was statistically indistinguishable ($F_{1,10} = 1.91$, $P > 0.05$) from that to untreated plants (Fig. 2B, a). Thus, female *H. virescens* are able to identify infested plants on the basis of chemical cues consisting of volatile compounds emitted by the plants at night.

Some of the volatiles in the nocturnal volatile profile were also present in the diurnal profile. However, others were released only at night. To determine the importance of volatiles produced at night, relative to those produced during the day, we repeated the behavioural assays using the diurnal blends. Although these daytime volatiles produced avoidance behaviour (Fig. 2B, c; $F_{1,10} = 55$, $P < 0.001$), they were significantly less repellent than the nocturnal volatiles (Fig. 2B, b), indicating that the moths are specifically repelled by the night-time volatile blends. To further examine the extent to which the repellence of the night-time volatiles is due to the presence of compounds produced exclusively at night, the experiment was repeated using a synthetic blend containing only these compounds. The presence of these exclusively nocturnal compounds alone was sufficient to explain the moth repellence effect (Fig. 2B, d; $F_{1,10} = 145$, $P < 0.0001$).

It is well established that induced plant volatiles function as signals between plants and the natural enemies of insect herbivores. The discovery that the herbivores themselves exploit the information present in night-time volatile blends to avoid oviposition on previously damaged plants reveals a new dimension of chemically

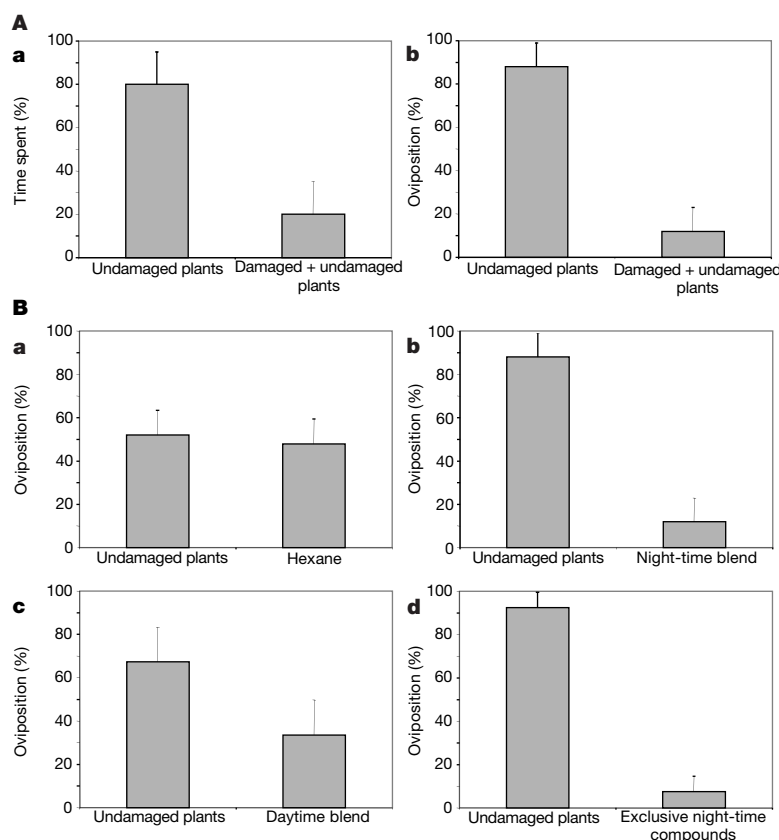


Figure 2 Response of female moths to plant volatiles. **A**, Nocturnal flight responses of three mated *H. virescens* females to tobacco plants. On one side of the cage, only undamaged plants (6) were used. On the other side, two plants that had been infested by ten third-instar *H. virescens* larvae were included. Larvae were removed from the plants before moths were introduced. **a**, Per cent of time during one hour of observation spent on each side of the cage, **b**, per cent of eggs oviposited on each side of the cage. **B**, Oviposition preference by *H. virescens* females for tobacco plants that had been treated

with synthetic blend. On one side of the cage only undamaged plants (6) were used, on the other side two plants were treated (synthetic blends on rubber septa) placed among four undamaged and untreated plants. **a**, Solvent (hexane) used as control. **b**, A nocturnal blend compared to undamaged plants. **c**, A diurnal blend compared to undamaged plants. **d**, A blend containing compounds released exclusively at night compared to undamaged plants.

mediated plant–insect interactions and raises a number of issues. The fitness advantages to herbivores of avoiding oviposition on induced plants are obvious, as such plants are likely to host not only larvae that represent potential competitors for the moth's offspring but also potentially a population of natural enemies attracted by the volatile blend²¹. Moreover, the associated induction of direct defence mechanisms means that infested plants are likely to contain chemical toxins and to be of lower nutritional value than uninfested plants^{22,23}. It is less clear whether plants benefit significantly from the release of nocturnal volatiles or whether such release is merely a physiological by-product associated with diurnal volatile production. The protection enjoyed by undamaged plants that reside near induced plants is interesting, but it is not certain whether this phenomenon would have been significant in the ancestral environment or represents only an effect of high-density agricultural cultivation. If plants do benefit from advertising their status to herbivores, it raises the question of whether herbivore-induced signals first evolved as parasitoid attractants or as herbivore repellents; alternatively, the dual functions of herbivore-induced plant signals may have evolved simultaneously.

The recognition that plant volatile signals, long known to be important in the mediation of plant–parasitoid interactions, also transmit information to herbivores expands our view of tritrophic systems and is significant with regard to our understanding of the selective pressures governing the evolution of such signals. We are currently exploring whether moths can interpret the herbivore-specific information that these signals convey to parasitoids¹³ or rather use these signals only as generalized indicators of insect feeding. If moths can interpret the higher-order information content of these signals, then they may be making sophisticated choices based on the likely presence of particular larval competitors and perhaps even of particular predators and parasitoids. □

Methods

Volatile collection and analysis

All experiments were conducted in an insect-free greenhouse (temperature $29 \pm 4^\circ\text{C}$). Ten third-instar larvae of either *H. virescens*, *H. zea* or *M. sexta* were allowed to feed continuously on the leaves of an eight-week-old, greenhouse-grown tobacco plant (*Nicotiana tabacum* strain K326) enclosed in a volatile collection chamber²⁴, beginning the night before collection commenced.

Plant volatiles were collected 24-h a day in 2-h intervals by pulling 1 l of the air passing over the plant (51 min^{-1}) through Super Q adsorbent (25 mg) traps at the base of the volatile-collection chamber; the remainder of the air vented out of the bottom of the system²⁴. Traps were rinsed with 150 μl methylene chloride, 400 ng of *n*-octane and nonyl acetate were added as internal standards, and samples were analysed by gas chromatography and mass spectrometry¹³ (electron impact and chemical ionization with isobutane reagent gas). Volatile compounds were identified by comparison of chromatographic retention times and mass spectra with those of commercially available standards analysed on the same instruments. Quantification was based on peak area (flame ionization detector) relative to that of internal standards.

Average volatile released by plants from 11:00 to 13:00, 50 h after initial damage: (ng h⁻¹, s.d.): (*E*)-2-hexenal (190.93, 23.06); (*Z*)-3-hexen-1-ol (62.04, 40.08); (*E*)- β -ocimene (5510.46, 789.72); indole (212.67, 45.03); β -caryophyllene (7481.88, 823.40); α -humulene (221.88, 13.08); (*E,E*)- α -farnesene (1220.48, 152.81), unidentified sesquiterpene (1148.25, 165.85).

Average volatile released by plants from 19:00 to 21:00, 58 h after initial damage: (ng h⁻¹, s.d.): (*E*)-2-hexenal (488.44, 36.78); (*Z*)-3-hexen-1-ol (792.70, 160.12); (*Z*)-3-hexenyl acetate (987.99, 97.12); (*E*)- β -ocimene (1610.46/349.32); (*Z*)-3-hexenyl isobutyrate (123.84, 23.76); (*Z*)-3-hexenyl butyrate (488.85, 121.67) indole (212.67, 45.03); (*Z*)-3-hexenyl tiglate (1328.87, 198.76); β -caryophyllene (2452.88, 223.46); α -humulene (83.48, 35.17); (*E,E*)- α -farnesene (888.88, 154.98), unidentified sesquiterpene (960.59, 200.91).

Behavioural assays

Twelve eight-week-old, potted, greenhouse-grown tobacco plants were placed in a screened cage ($4 \times 2 \times 3\text{ m}$) with six plants on each side of the cage (80 cm between plants). On one side only undamaged plants were used, on the other side two of the plants had been fed on by 10 laboratory-reared third-instar *H. virescens* (two larvae per leaf) for 48 h. Larvae were removed from plants before moth release. Three mated *H. virescens* females (18 females per treatment) were released in the central sector of the cage at dusk. To account for a possible directional preference by the moths, two cages were used so in one the damaged plants were on the north end and in the other they were on the south end of the cage. The moths' behaviour was visually observed for one hour (from approximately

19:00 to 20:00) and the time (min.) spent by moths on each side of the cage was recorded. Egg numbers per plant on each side of the cage were counted at 06:00 the next morning. Similar procedures were used for assays measuring oviposition and behaviour of female moths in response to plants with or adjacent to synthetic blends, but in this case plants on both sides of the cages were undamaged. For these treatments, pots of undamaged plants with volatile-releasing rubber septa on wooden sticks (three per plant) replaced the damaged plants. We evaluated the response of moths to three synthetic volatile blends (nocturnal volatiles, exclusively nocturnal volatiles, and diurnal volatiles). Rubber septa without the synthetic blend were placed on the control side to neutralize any visual preference. Each bioassay was conducted on three days (two repetitions each) to account for day-to-day variation. Preference between the undamaged and damaged plants by the herbivores was subjected to analysis of variance (an arc-sine square root transformation of the percentage data was used; SAS Institute, Cary, North Carolina).

Synthetic blends

Synthetic blends were formulated according to a method developed to predict the release ratio of volatile compounds from rubber septa^{19,20}. Calculations of predicted release ratios are based on relative vapour pressures¹⁹ of the components and the original amounts released in the natural blend. Blends were dissolved in hexane and 0.3 ml of a blend solution was pipetted onto a rubber septum. Volatiles released by the blend when formulated on rubber septa were sampled and analysed. The relative amounts were adjusted to correct for slight deviations from the predicted amount. Compounds used to prepare synthetic blends were obtained from Sigma-Aldrich or from Chemical Samples and analysed by gas chromatography and mass spectrometry to determine purity. All synthetic compounds were at least 98% pure except for ocimene where both isomers were present (60% *trans* and 40% *cis*).

Release rates of synthetic blends: daytime blend 1, (*E*)-2-hexenal (200 ng h⁻¹); 2, (*E*)- β -ocimene (5,500 ng h⁻¹); 3, β -caryophyllene (7,300 ng h⁻¹); 4, α -humulene (200 ng h⁻¹); 5, (*E,E*)- α -farnesene (1,180 ng h⁻¹). Night-time blend 1, (*E*)-2-hexenal (500 ng h⁻¹); 2, (*Z*)-3-hexen-1-ol (850 ng h⁻¹); 3, (*Z*)-3-hexenyl acetate (1,000 ng h⁻¹); 4, (*E*)- β -ocimene (1,500 ng h⁻¹); 5, (*Z*)-3-hexenyl isobutyrate (100 ng h⁻¹); 6, (*Z*)-3-hexenyl butyrate (500 ng h⁻¹); 7, (*Z*)-3-hexenyl tiglate (1,200 ng h⁻¹); 8, β -caryophyllene (2,200 ng h⁻¹); 9, (*E,E*)- α -farnesene (900 ng h⁻¹). Exclusive night-time blend 1, (*Z*)-3-hexen-1-ol (850 ng h⁻¹); 2, (*Z*)-3-hexenyl acetate (1,000 ng h⁻¹); 3, (*Z*)-3-hexenyl isobutyrate (100 ng h⁻¹); 4, (*Z*)-3-hexenyl butyrate (500 ng h⁻¹); 5, (*Z*)-3-hexenyl tiglate (1,200 ng h⁻¹).

Received 19 December 2000; accepted 31 January 2001.

- Dicke, M. Local and systemic production of volatile herbivore-induced terpenoids: their role in plant–carnivore mutualism. *J. Plant Physiol.* **143**, 465–472 (1994).
- Turlings, T. C. J., Tumlinson, J. H. & Lewis, W. J. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* **250**, 1251–1253 (1990).
- McCall, P. J., Turlings, T. C. J., Lewis, W. J. & Tumlinson, J. H. Role of plant volatiles in host location by the specialist parasitoid *Microplitis croceipes* Cresson (Braconidae:Hymenoptera). *J. Insect Behav.* **6**, 625–639 (1993).
- Pallini, A., Jansen, A. & Sabelis, M. Odour-mediated responses of phytophagous mites to conspecific and heterospecific competitors. *Oecologia* **110**, 179–185 (1997).
- Karban, R. & Baldwin, I. T. *Induced Responses to Herbivory* (Univ. Chicago Press, Chicago, Illinois, 1997).
- Paré, P. W. & Tumlinson, J. H. Induced synthesis of plant volatiles. *Nature* **385**, 30–31 (1997).
- Loughrin, J. H., Manukian, A., Heath, R. R. & Tumlinson, J. H. Diurnal cycle of emission of induced volatile terpenoids by herbivore-injured cotton plants. *J. Chem. Ecol.* **21**, 1217–1227 (1994).
- Turlings, T. C. J. et al. How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proc. Natl Acad. Sci. USA* **92**, 4169–4174 (1995).
- Karban, R. Resistance against spider mites in cotton induced by mechanical abrasion. *Entomol. Exp. Appl.* **37**, 137–141 (1985).
- Dicke, M. et al. Isolation and identification of volatile kairomone that affects acarine predator–prey interactions. Involvement of host plant in its production. *J. Chem. Ecol.* **16**, 381–396 (1990).
- Mattiacci, L., Dicke, M. & Posthumus, M. A. Induction of parasitoid attracting synomone in brussels sprouts plants by feeding of *Pieris brassicae* larvae: Role of mechanical damage and herbivore elicitor. *J. Chem. Ecol.* **20**, 2229–2247 (1994).
- Alborn, H. T. et al. An elicitor of plant volatiles from beet armyworm oral secretion. *Science* **276**, 945–948 (1997).
- De Moraes, C. M. et al. Herbivore-infested plants selectively attract parasitoids. *Nature* **393**, 570–573 (1998).
- Dicke, M. Volatile spider-mite pheromone and host plant kairomone, involved in spaced-out gregariousness in spider mite *Tetranychus urticae*. *Physiol. Entomol.* **11**, 251–262 (1986).
- Schultz, S. et al. Host plant selection of the Colorado Potato Beetle as influenced by damage induced volatiles of the potato plant. *Naturwissenschaften* **84**, 212–217 (1997).
- Harari, A. R., Ben-Yakir, D. & Rosen, D. Mechanisms of aggregation behavior in *Maladera matrida* Argaman (Coleoptera: Scarabaeidae). *J. Chem. Ecol.* **20**, 361–371 (1994).
- Landolt, P. J., Tumlinson, J. H. & Alborn, D. H. Attraction of Colorado potato beetle (Coleoptera: Chrysomelidae) to damaged and chemically induced potato plants. *Environ. Entomol.* **28**, 973–978 (1999).
- Landolt, P. J. Effects of host plant leaf damage on cabbage looper moth attraction and oviposition. *Entomol. Exp. Appl.* **67**, 79–85 (1993).
- Heath, R. R. & Tumlinson, J. H. Prediction of release ratios of multicomponent pheromones from rubber septa. *J. Chem. Ecol.* **12**, 2081–2088 (1986).
- Heath, R. R., Teal, P. E. A., Tumlinson, J. H. & Mengelkoch, L. J. Correlation of retention times on liquid crystal capillary column with reported vapor pressures and half-lives of compounds used in pheromone formulations. *J. Chem. Ecol.* **12**, 2133–2143 (1986).
- Thaler, J. S. Jasmonate-inducible plant defenses cause increased parasitism of herbivores. *Nature* **399**, 686–688 (1999).

22. Agrawal, A. A. Induced responses to herbivory and increased plant performance. *Science* **279**, 1201–1202 (1998).
23. Coley, P. D., Bryant, J. P. & Chapin, F. S. Resource availability and plant antiherbivore defense. *Science* **230**, 895–899 (1985).
24. Heath, R. R. & Manukian, A. An automated-system for use in collecting volatile chemicals released from plants. *J. Chem. Ecol.* **20**, 593–608 (1994).

Acknowledgements

We thank G. W. G. De Moraes for discussions; H. T. Alborn, J. G. Hildebrand, P. J. Landolt, W. J. Lewis, K. G. Ross and J. R. Ruberson for comments on the manuscript; and B. Dueben and M. Sammons for technical assistance.

Correspondence and requests for materials should be addressed to J.H.T. (e-mail: jtumlinson@gainesville.usda.ufl.edu).

Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia

Jason B. Mattingley*, Anina N. Rich*, Greg Yelland† & John L. Bradshaw†

* School of Behavioural Science, University of Melbourne, Victoria 3010, Australia
† Department of Psychology, Monash University, Victoria 3800, Australia

Synaesthesia is an unusual perceptual phenomenon in which events in one sensory modality induce vivid sensations in another^{1,2}. Individuals may 'taste' shapes³, 'hear' colours⁴, or 'feel' sounds⁵. Synaesthesia was first described over a century ago⁶, but little is known about its underlying causes or its effects on cognition. Most reports have been anecdotal or have focused on isolated unusual cases^{3,7–9}. Here we report an investigation of 15 individuals with colour-graphemic synaesthesia, each of whom experiences idiosyncratic but highly consistent colours for letters and digits. Using a colour–form interference paradigm, we show that induced synaesthetic experiences cannot be consciously suppressed even when detrimental to task performance. In contrast, if letters and digits are presented briefly and masked, so that they are processed but unavailable for overt report, the synaesthesia is eliminated. These results show that synaesthetic experiences can be prevented despite substantial processing of the sensory stimuli that otherwise trigger them. We conclude that automatic binding of colour and alphanumeric form in synaesthesia arises after initial processes of letter and digit recognition are complete.

We studied 15 individuals with colour-graphemic synaesthesia and 15 non-synaesthetic controls. Each synaesthete reported vivid and immediate sensations of colour for specific letters and digits. All reported having had synaesthesia since childhood, and many had biological relatives with the phenomenon, consistent with previous reports¹⁰. Our study focused on colour-graphemic synaesthesia because it is the most common form¹⁰, and because it has received considerable attention¹¹.

A test of consistency verified the presence of synaesthesia in our group⁴. Participants were each given a 150-item list containing letters (A–Z), digits (0–9) and words. They described their synaesthetic colour for each item (or an arbitrary colour in the case of non-synaesthetic controls). Three months later, without warning, the synaesthetes were given the same list and again asked to indicate their synaesthetic colour for each item. For controls, the retest was given just one month later, thus giving them a potential advantage. The synaesthetes were highly consistent in their responses overall, significantly more so than the controls ($F_{1,28} = 162.56$, $P < 0.0001$; Fig. 1). These findings show that the unusual sensations experienced

by synaesthetes, although idiosyncratic, remain highly stable over time⁴.

Colour-graphemic synaesthesia may be triggered by automatic co-activation of independent brain areas responsible for processing colour and symbolic form^{12,13}. This fits with synaesthetes' subjective accounts of the involuntary nature of their experiences. We therefore manipulated the physical colours of alphanumeric characters so that they differed from the synaesthetic colours induced. Our approach was based on the Stroop effect¹⁴, in which naming the print colour of an incongruent colour word (for example, RED printed in blue) takes significantly longer than naming the print colour of a congruent colour word (for example, RED printed in red), or a colour patch. This slowing reflects interference arising from an involuntary word-reading response¹⁵.

We began by comparing synaesthetes and controls on the standard Stroop task to check for any baseline differences in their susceptibility to interference. Colour words were displayed in either congruent or incongruent colours; solid colour patches were used in a baseline condition (see Methods). Participants named the colour of each stimulus aloud. As expected, both groups were significantly slower to name colours in the incongruent than the congruent condition ($F_{1,28} = 119.72$, $P < 0.001$). Critically, there was no overall difference between the groups ($F_{1,28} = 0.66$, $P > 0.10$) and no interaction ($F_{1,28} = 1.56$, $P > 0.10$; Fig. 2a), indicating equivalent interference for synaesthetes and controls. Moreover, their colour naming times in the baseline condition were the same (535 ms versus 536 ms; $t_{28} = 0.03$, $P > 0.10$).

If synaesthesia is an involuntary phenomenon, then having participants judge the physical colour of an alphanumeric character that elicits an incongruent synaesthetic colour should yield significant interference, and slow response times accordingly. We therefore constructed unique stimulus ensembles for each synaesthete, which

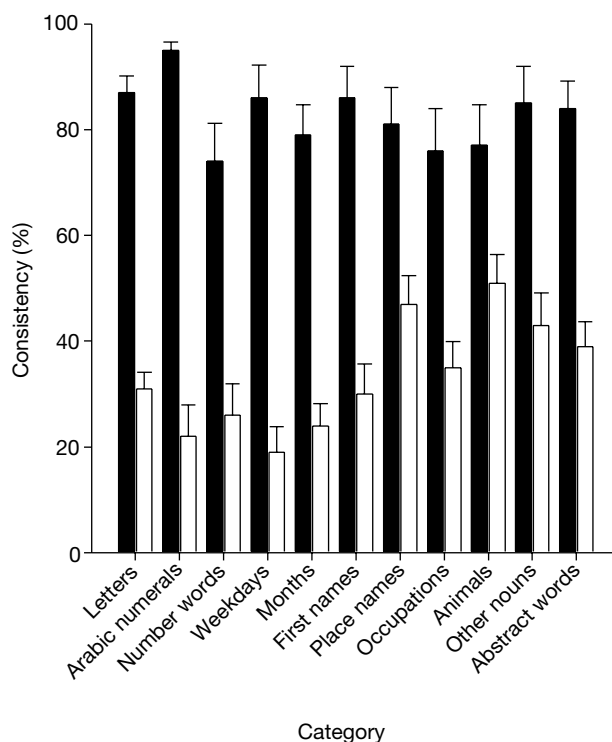


Figure 1 Mean (+1 s.e.) consistency of colour associations for 150 items (letters, arabic numerals and words), plotted separately for each of 11 categories tested. Results for synaesthetes (filled bars) represent performance with a 3-month retest interval; those for non-synaesthetic controls (open bars) represent performance with a 1-month retest interval. For every category tested synaesthetes were more consistent in their colour associations than non-synaesthetic controls.

were based on a detailed colour-matching procedure, to obtain congruent and incongruent stimulus sets (see Methods).

Participants named aloud the physical colour of alphanumeric stimuli presented individually on a computer display. In one experiment, synaesthetically congruent and incongruent stimuli were presented in separate blocks of trials. Synaesthetes took significantly longer to name display colours when these were incongruent with their induced synaesthetic colours, as compared with the congruent condition in which the physical and synaesthetic colours were matched ($F_{1,28} = 12.00$, $P < 0.01$; Fig. 2b). Notably, there was no difference between congruent and incongruent conditions for the non-synaesthetic controls ($F_{1,28} = 0.31$, $P > 0.10$; Fig. 2b).

In a second experiment, congruent and incongruent trials were randomly intermingled. Additional 'neutral' stimuli, comprising non-alphanumeric typographical symbols, were also included (see Methods). Synaesthetes again showed a significant effect of congruency ($F_{2,28} = 7.43$, $P < 0.01$; Fig. 2c), whereas the controls did not ($F_{2,28} = 1.17$, $P > 0.10$). These results show that colour-graphemic synaesthetes cannot consciously suppress their unusual colour sensations. This is consistent with individual case reports of

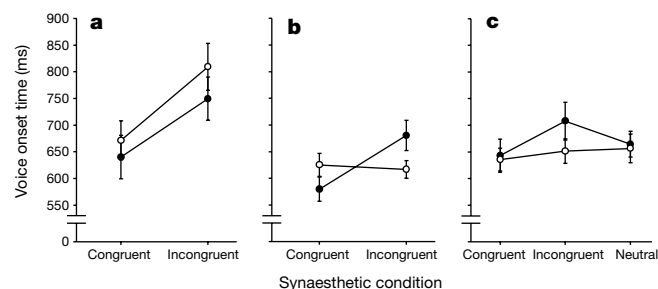


Figure 2 Mean voice-onset times (± 1 s.e.) for colour naming in the standard and synaesthetic Stroop tasks, plotted as a function of congruency condition. Data are shown as separate lines for synaesthetes (solid symbols) and non-synaesthetic controls (open symbols). **a**, Randomized presentation of congruent and incongruent colour words in the standard Stroop task. Means for baseline condition are reported in the text. **b**, Blocked presentation of synaesthetically congruent and incongruent alphanumeric characters. **c**, Randomized presentation of synaesthetically congruent, incongruent and neutral alphanumeric characters. The faster responses of synaesthetes relative to controls in the congruent condition of **b** (but not **c**) implies a strategic benefit from blocked presentation.

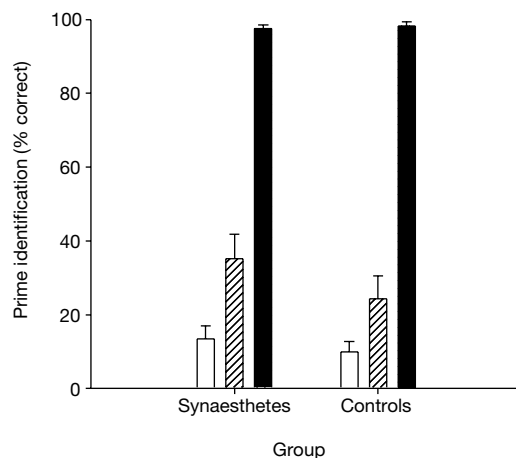


Figure 3 Mean percentage correct prime identification (± 1 s.e.) for the priming experiments. Data are plotted separately for synaesthetes and non-synaesthetic controls. Prime durations were 28 ms (open bars), 56 ms (hatched bars) and 500 ms (filled bars); identification scores were not different for the two groups at any of the durations ($P > 0.10$ for all comparisons).

synaesthetic interference^{7-9,11}, as well as our synaesthetes' subjective experiences of overwhelming mental effort in the synaesthetically incongruent trials.

We next determined whether the involuntary binding of colour and form occurs when the triggering stimuli are not available for conscious report. Visual stimuli that are displayed briefly and masked may receive considerable unconscious processing¹⁶⁻¹⁸. We modified our synaesthetic Stroop task so that an achromatic alphanumeric character was presented briefly and masked. Participants named the colour of a target patch presented immediately after the masked prime. The colour of the target patch was congruent, incongruent or neutral with respect to the synaesthetic colour induced by the alphanumeric prime. In an initial experiment, the prime was presented for 500 ms so that it was clearly visible. We tested prime identification in a separate task in which participants had to name the alphanumeric prime, instead of naming the colour of the target patch (see Fig. 3).

In the colour-naming task, synaesthetes were significantly affected by prime-target congruency ($F_{2,28} = 14.22$, $P < 0.001$; Fig. 4a). Naming times were slower in the incongruent and neutral conditions than in the congruent condition ($P < 0.01$ for both). In contrast, controls showed no significant effect of congruency ($F_{2,28} = 0.137$, $P > 0.10$; Fig. 4b). The unexpectedly slow naming times in the neutral condition reveal a further intriguing characteristic of the synaesthetic group (see Fig. 4a), most of whom reported 'weak' or 'faint' colours for these non-alphanumeric symbols. The achromatic primes in the neutral condition thus elicited incongruent colours for most of the synaesthetes, leading to significant interference.

In a further experiment, primes were presented for either 56 or 28 ms, in separate blocks of trials. At these durations, conscious identification of the alphanumeric primes was severely curtailed (Fig. 3); indeed, participants were not aware of their presence. As shown in Fig. 4a, the effect of congruency on colour-naming times at these shorter prime durations was completely absent in the

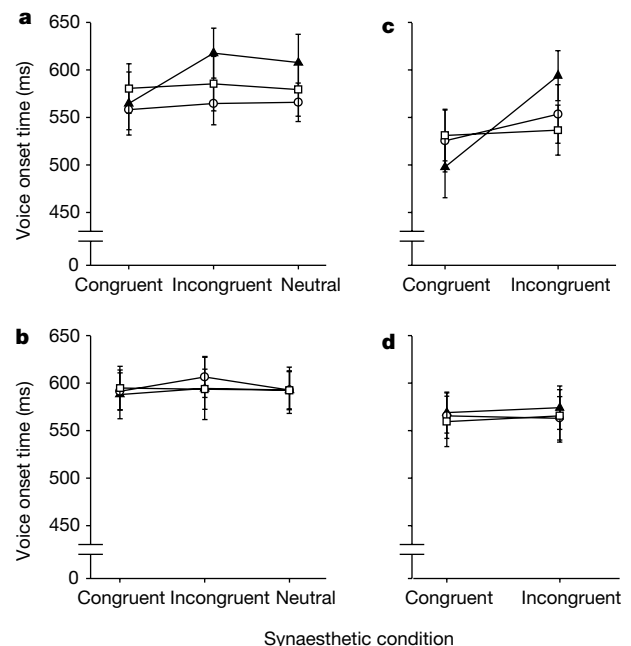


Figure 4 Mean voice-onset times (± 1 s.e.) for colour naming in the visible priming (500 ms) and masked priming (56 and 28 ms) experiments, plotted as a function of synaesthetic congruency condition. **a**, Synaesthetes: standard priming. **b**, Controls: standard priming. **c**, Synaesthetes: temporal-gap priming. **d**, Controls: temporal-gap priming. Data are shown as separate lines for each of the three prime durations of 500 ms (filled triangles), 56 ms (open circles) and 28 ms (open squares).

synaesthetes, whose performance was now similar to that of the controls ($P > 0.10$; Fig. 4b). These results show for the first time, to our knowledge, that synaesthesia is eliminated when inducing stimuli are unavailable for conscious report.

We verified that the primes were processed unconsciously in a separate interference task involving letter naming. In this task letters replaced the colour-patch targets, and were either congruent (such as $a \rightarrow A$) or incongruent (such as $b \rightarrow A$) with the preceding primes. At prime durations of 56 and 28 ms, both groups were significantly slower to name letters in the incongruent condition (costs of 21 ms and 7 ms, respectively; both $P < 0.001$). This interference effect indicates unconscious processing of the letter prime, consistent with previous reports for whole-word priming¹⁷.

To determine whether the absence of interference might be due to the time required for synaesthetic colours to be elicited by the primes, we repeated the initial priming experiments with the prime-target onset interval held constant (see Methods). Alphanumeric primes appeared for 28 or 56 ms, and were followed by a 100-ms backward mask to curtail identification. The screen then remained blank for 372 or 344 ms before the onset of the coloured target. A 500-ms prime duration was also included (no prime-target gap), in which all primes were clearly visible. Thus, the onset asynchrony between prime and target was held constant at 500 ms in all three conditions. At the 500-ms prime duration synaesthetes' colour naming was significantly slower for incongruent versus congruent conditions ($F_{1,12} = 15.12$, $P < 0.01$; Fig. 4c), whereas there was no such effect for controls ($F_{1,12} = 0.45$, $P > 0.10$; Fig. 4d). This replicates the interference effect observed in the initial visible priming experiment (see Fig. 4a). With masked priming, however, the congruency effect disappeared entirely for synaesthetes ($F_{1,12} = 2.06$, $P > 0.10$; Fig. 4c), and was again absent for controls ($F_{1,12} = 3.25$, $P = 0.10$; Fig. 4d) despite the constant interval between prime and target across all conditions.

Our findings have important implications for understanding the automatic binding of colour and form in synaesthesia. They suggest that synaesthetic colours cannot be consciously suppressed, and that robust cognitive interference occurs when there is a conflict between real and synaesthetic colours. Our results go beyond previous single-case reports^{7–9,11} by showing that such interference is a ubiquitous feature of colour-graphemic synaesthesia, thus providing an objective, cognitive marker for the phenomenon. They also reveal that the obligatory binding of colour and form in synaesthesia can be broken when inducing stimuli are masked, rendering them unavailable for conscious report. We conclude that synaesthetic interactions arise after initial processing of visual form, and that overt recognition of inducing stimuli is crucial. Our view is consistent with the idea that synaesthesia is elicited by selectively attended stimuli that are available for conscious report¹⁹. □

Methods

Subjects

Fifteen colour-graphemic synaesthetes (13 female, 13 right-handed; mean age = 41.26 years; range = 18–60 yr) and 15 non-synaesthetic controls (matched for age, sex and handedness) were tested. All participants were screened for neurological and colour vision impairments. Synaesthetes completed a questionnaire that documented their unusual experiences and personal history.

Synaesthetic Stroop tasks

Individual synaesthetes selected from a standard palette the colour that best matched their synaesthetic experience for each item in a set of alphanumeric characters (letters A–Z, digits 0–9). They rated their choices on a 5-point scale. The six characters yielding the best colour matches for each synaesthete were used in the experimental tasks. The mean colour-match rating for the group was 4.0 ('very well matched'). These synaesthetically matched characters were the 'congruent' stimuli; 'incongruent' stimuli were obtained by making the same characters different in colour to those elicited synaesthetically.

Participants were tested individually in a sound attenuated booth. Their voice-onset times were recorded by a Sennheiser HMD224 head-mounted microphone interfaced with a PC (IBM compatible 486) that controlled all aspects of stimulus presentation and response recording (DMTG software; K. Forster and J. Forster, Monash Univ. and Univ.

Arizona). Stimuli appeared individually in the centre of a 38-cm monitor (Phillips Brilliance 15A) against a black background. Trials were self-paced by means of a foot pedal.

In the standard Stroop task, stimuli were uppercase colour words (BLUE, GREEN, PURPLE, RED, WHITE, YELLOW) presented in Times Roman font ($1.4^\circ \times 2.6\text{--}5.4^\circ$ at a viewing distance of 50 cm). Each colour word appeared eight times, four in its congruent colour and four in a single incongruent colour, for a total of 48 trials per block. Stimuli in the baseline condition consisted of solid rectangles ($3.4^\circ \times 4.6^\circ$) in one of the six colours, each of which appeared eight times for a total of 48 trials in a block. In the modified Stroop task, synaesthetically congruent and incongruent alphanumeric characters (Times Roman, $1.4^\circ \times 0.8^\circ$) were presented either in separate blocks of 36 trials, or randomly intermingled together with non-alphanumeric symbols such as an asterisk ('neutral' condition) within a single block of 48 trials. In the randomized presentation, congruent, neutral and incongruent stimuli appeared in a ratio of 1:1:2, with each colour being equiprobable. Each block was preceded by five practice trials. In all tasks, participants named aloud the colour of the visual target (word, character or colour patch). Stimuli remained visible for 4,000 ms or until a response was made.

Priming experiments

Participants named the colour of target patches consisting of six overlapped typographical symbols (Times Roman, $1.4^\circ \times 1.0^\circ$). These targets formed a pattern mask for the preceding primes, which were the same alphanumeric characters as those used in the synaesthetic Stroop tasks, presented in light grey. Each trial commenced with a light grey forward mask ($1.7^\circ \times 3.3^\circ$), followed by a prime of variable duration (28, 56 or 500 ms in separate blocks), and then a coloured target which remained visible for 4,000 ms or until a response was made. Targets appeared in one of the six colours used in the synaesthetic Stroop tasks. Congruent, neutral and incongruent trials were randomly intermingled within a single block of 48 trials, in a ratio of 1:1:2. In the 'temporal gap' variant of the priming experiments, conducted with 13 of the 15 synaesthetes, only congruent and incongruent trials were included (1:1, 48 trials per condition). Masked primes were followed immediately by an achromatic pattern mask for 100 ms, a blank interval of either 372 or 344 ms, and then a coloured target which remained visible for 4,000 ms or until a response was made.

In a separate task, participants' ability to identify the primes was assessed. Stimuli and display sequences were identical to those used in the colour-naming task. Participants were instructed to identify the primes rather than name target colours. Their untimed responses were recorded verbatim. There were 48 trials, with equal numbers of stimuli from each of the three prime durations. To verify that primes were processed at durations of 28 and 56 ms, a letter-to-letter priming task was conducted in which participants named uppercase target letters aloud. These were preceded by lowercase primes. Primes could appear for 28, 56 or 500 ms, in separate blocks of 48 trials. There were equal numbers of congruent (for example, $a \rightarrow A$) and incongruent (for example, $b \rightarrow A$) trials. All other aspects of the display sequence were the same as those used in the synaesthetic priming experiments.

Received 8 November; accepted 12 December 2000.

1. Baron-Cohen, S. & Harrison, J. E. (eds) *Synaesthesia: Classic and Contemporary Readings* (Blackwell, Cambridge, 1997).
2. Cytowic, R. E. *Synaesthesia: A Union of the Senses* (Springer, New York, 1989).
3. Cytowic, R. E. *The Man Who Tasted Shapes* (Abacus, London, 1993).
4. Baron-Cohen, S., Wyke, M. A. & Binnie, C. Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. *Perception* **16**, 761–767 (1987).
5. Luria, A. R. *The Mind of a Mnemonist* (Basic Books, London, 1969).
6. Galton, F. Visualized numerals. *Nature* **21**, 252–256 (1880).
7. Mills, C. B., Boteler, E. H. & Oliver, G. K. Digit synaesthesia: A case study using a Stroop-type test. *Cog. Neuropsychol.* **16**, 181–191 (1999).
8. Odgaard, E. C., Flowers, J. H. & Bradman, H. L. An investigation of the cognitive and perceptual dynamics of a colour-digit synaesthete. *Perception* **28**, 651–664 (1999).
9. Wollen, K. A. & Ruggiero, F. T. Colored-letter synaesthesia. *J. Mental Imagery* **83**–86 (1983).
10. Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J. & Bolton, P. Synaesthesia: prevalence and familiarity. *Perception* **25**, 1073–1079 (1996).
11. Dixon, M. J., Smilek, D., Cudahy, C. & Merikle, P. M. Five plus two equals yellow. *Nature* **406**, 365 (2000).
12. Paulesu, E. et al. The physiology of coloured hearing: A PET activation study of colour-word synaesthesia. *Brain* **118**, 661–676 (1995).
13. Baron-Cohen, S., Harrison, J., Goldstein, L. H. & Wyke, M. Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception* **22**, 419–426 (1993).
14. Stroop, J. R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935).
15. MacLeod, C. M. Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* **109**, 163–203 (1991).
16. Dehaene, S. et al. Imaging unconscious semantic priming. *Nature* **395**, 597–600 (1998).
17. Forster, K. I. & Davis, C. Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cog.* **10**, 680–698 (1984).
18. Marcel, A. J. Conscious and unconscious perception: experiments on visual masking and word recognition. *Cog. Psychol.* **15**, 197–237 (1983).
19. Grossenbacher, P. G. in *Synaesthesia: Classic and Contemporary Readings* (eds Baron-Cohen, S. & Harrison, J. E.) (Blackwell, Cambridge, 1997).

Acknowledgements

We wish to thank A. Kritikos, M. O'Boyle, P. Wilken and M. Williams for their comments on the manuscript. This work was supported by a grant from the Australian Research Council to J.B.M.

Correspondence and requests for materials should be addressed to J.B.M. (e-mail: j.mattingley@psych.unimelb.edu.au).

Vertical interactions across ten parallel, stacked representations in the mammalian retina

Botond Roska & Frank Werblin

Department of Molecular and Cell Biology, University of California at Berkeley, 145 LSA, Berkeley, California 94720, USA

The mammalian visual system analyses the world through a set of separate spatio-temporal channels^{1,2}. The organization of these channels begins in the retina^{1,3}, where the precise laminations of both the axon terminals of bipolar cells and the dendritic arborizations of ganglion cells suggests the presence of a vertical stack of neural strata at the inner plexiform layer (IPL)^{3–11}. Conversely, many inhibitory amacrine cell classes are multiply or diffusely stratified¹², indicating that they might convey information between strata. On the basis of the diverse stratification and physiological properties of ganglion cells, it was suggested that the IPL contains a parallel set of representations of the visual world^{3,7} embodied in the strata and conveyed to higher centres by the classes of ganglion cells whose dendrites ramify at that stratum. Here we show that each stratum receives unique and substantively different excitatory and inhibitory neural inputs that are integrated to form at least ten different, parallel space-time spiking outputs. The response properties of these strata are ordered in the time domain. Inhibition through GABA_C receptors extracts spatial edges in neural representations and seems to separate the functional properties of the strata. We describe a new form of neuronal interaction that we call 'vertical inhibition' that acts not laterally, but between strata.

We studied ON and OFF ganglion cells with concentric receptive fields, because the dendrites of many of these ganglion cells ramify in a single, narrow stratum within the IPL. This allowed us to 'read out' activity at specific strata by recording from anatomically identified ganglion cells. First, we patch-clamped ganglion cells in the whole-mount rabbit retina in the loose cell-attached mode, and measured spikes (output). After the spiking measurements, we recorded from the same cells with a second electrode in the whole-cell configuration and measured excitatory currents (currents that reflect the input from bipolar cells) and inhibitory currents (currents that reflect the feedforward input from amacrine cells) (Fig. 1a). The inhibitory and excitatory inputs could be separately measured by voltage clamping the cells at appropriate potentials (see Methods). To get a sensitive measure of space-time activity we took measurements from a single cell and stepped the visual stimulus, a 600- μ m flashed square (2–10 times bigger than the dendritic diameter of most of the measured cells), through the retina in one spatial dimension. The measured responses corresponding to each stimulus presentation were arranged¹³ according to the spatial position of the stimulus relative to the measured cell, thereby generating a space-time plot (Fig. 1d–g). This plot is an estimate of the activity pattern that this specific visual image would generate across a population of a given type of ganglion cells. This view is supported by the notion that each ganglion cell type covers and tiles¹⁴ the retina, although not necessarily in the same density as in our plots. Then we correlated the different excitatory, inhibitory and spiking patterns with the level of dendritic stratification, determined by intracellular staining and confocal reconstruction (Fig. 1b, c).

Individual members of ganglion cell types, defined by responses to different sizes of spots and by level of dendritic stratification (Table 1), show similar spiking patterns (Fig. 2). Figure 3 shows the correlation between depth within the IPL (Fig. 3a) and the

measured activity patterns (Fig. 3b) for seven of out of ten different classes of ON and OFF ganglion cells. Comparing the spiking patterns (Fig. 3b, column 1) shows that the seven classes extract different spatio-temporal components of the visual input. OFF cells, which ramify in the distal part of the IPL, characteristically spike outside the spatial boundaries of the stimulus (white lines in the space-time plots) at light ON and within the spatial boundaries of the stimulus at light OFF. ON cells, which ramify in the proximal part of the IPL, display a complementary pattern: they spike within the spatial boundaries of the stimulus at light ON and outside the spatial boundaries of the stimulus at light OFF. The spiking at light OFF is sometimes very weak or missing but the excitatory input is always present. The different spiking patterns represent different aspects of the visual stimulus. For example, cell type 3 displays spikes in a very narrow region of space outside the edges of the stimulus during the light flash, and inside the edges of the stimulus after the light flash. Cell type 7 displays a similar 'edge enhancement' but only transiently within the boundaries of the stimulus. In contrast, ganglion cell types 4 and 5 spike across the full dimension of the square, but only briefly at light OFF (type 4) or ON (type 5).

Comparing the spiking patterns (Fig. 3b, column 1) with classical receptive-field measurements (Fig. 3c, red traces) shows that some aspects of the spiking patterns are predictable from these measurements, whereas others are not. Strong surround inhibition, determined by comparing the spike frequency responses to different sizes

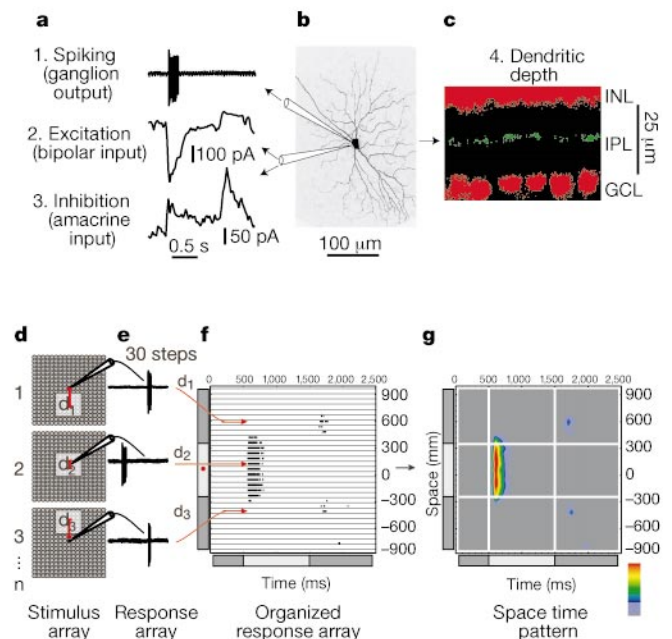


Figure 1 Space-time pattern measurements. **a–c**, Correlating activity with strata.

a, From each ganglion cell the spiking (output), excitatory currents and inhibitory currents (dendritic inputs) were measured with two electrodes in different patch-clamp configurations (see Methods). **b**, Stained ganglion cell. **c**, Ten-micrometre optical slice; cell nuclei (red) and ganglion dendrites (green) are indicated. INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer. **d–g**, Space-time pattern measurements. **d**, One-second flash of 600- μ m square stimulus presented every 10 s sequentially at 30 or 60 different locations (separated by 60 or 30 μ m) along a one-dimensional path. In the second dimension the square was centred on the measured cell. **e**, Spiking responses for three different stimulus presentations. **f**, Array of responses for 30 locations relative to the position of the stimulus centre (red spot). The y axis represents distance from the measuring electrode to the stimulus centre. **g**, Colour-coded space-time plot. Colours represent the spike frequency binned in time, the inhibitory current measurements binned in current, and the excitatory current measurements binned in minus current. Colour code (right corner), red indicates highest activity. The time and space dimensions of the stimulus are indicated by white bars.

of flashed squares, results in more activity around the edges and less activity in the centre of the square (cell types 3 and 7). The time course of activity around the edges of the stimulus is, however, not predictable: the spiking responses for both cell type 3 and 7 (OFF and ON response, respectively) for small flashed squares are sustained, and become more transient for larger squares, but the spiking activity within the boundaries of the stimulus is sustained for cell type 3 (light OFF) and transient for cell type 7 (at light ON). Moreover, the temporal and spatial characteristics of spiking patterns outside the stimulus boundary are not predictable for any of the cell classes from the receptive-field measurements (Fig. 3c).

In the space domain different ganglion cells show a transition from representing only the edges to representing a full but blurred version of the flashed square (Fig. 4a). In the mammalian retina there are only two classes of horizontal cell¹⁵ but many different classes of amacrine cell¹², which suggests that differential lateral feedback from amacrine cells to bipolar terminals^{16,17} is more likely to generate the different amounts of edge extraction. GABA_C (γ -aminobutyric acid) receptors^{18,19} are expressed mostly in bipolar axon terminals in the rabbit retina²⁰, so we tested whether blocking these receptors would eliminate edge enhancement. Figure 4b shows that blocking the GABA_C component of the feedback pathway for one of the most common ON ganglion cell types (cell type 7) eliminated its ability to extract edges ($n = 10$), thereby converting its response to a profile more like that of cell type 5 (Fig. 4a). These observations are consistent with the view that diversity in spatial representations is generated by differential lateral feedback to bipolar terminals.

The time course of responses was ordered with the level of stratification^{9,21} when ganglion-cell responses were tested with a small spot to minimize the effect of lateral interactions. Activity was more transient for ganglion cells ramifying in strata near the border between the ON and OFF ganglion-cell dendrites, and more sustained toward the proximal and distal regions of the IPL (Fig. 4c, d). Notably, with the exception of cell type 3, most of the sustained cells (cell types 1, 7, 8 and 9) also had a large, transient spiking response component (Fig. 3c). Blocking GABA and glycine receptors did not convert the transient responses to sustained responses (Fig. 4e; $n = 10$), suggesting that the response time

course is an inherent property of the photoreceptor–bipolar cell–ganglion cell pathway, and that amacrine inhibition is not the primary factor determining timing of the response for a small spot.

We evaluated the function of postsynaptic inhibition in the formation of the diverse spiking patterns by comparing the excitatory and inhibitory input patterns with the spiking patterns

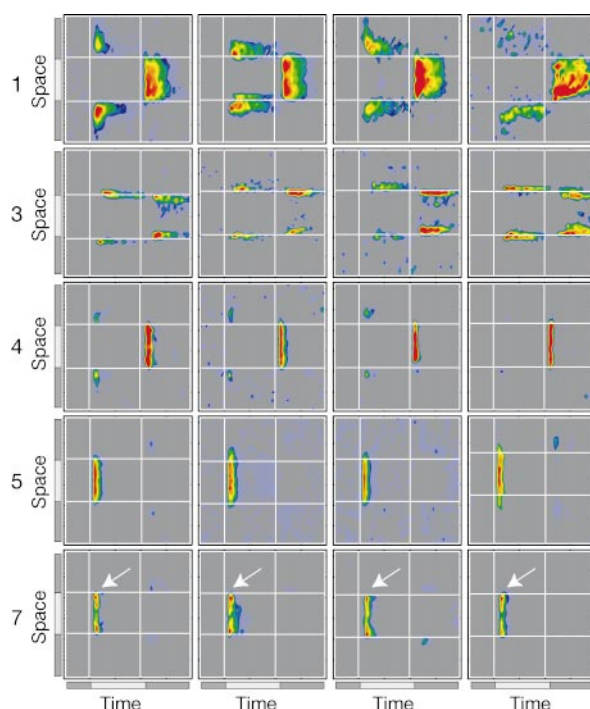


Figure 2 Spiking patterns in each row (measured from four different members of the same cell class) for five different classes of ganglion cells. The numbers at the beginning of each row refer to the classification shown in Table 1. The arrows in row 5 (cell type 7) point to the enhanced activity at the stimulus edge, a feature that is characteristic of this class. The units for space and time are the same as in Fig. 1g.

Table 1 Ganglion cell types, defined by response to spot sizes and level of dendritic stratification

Cell type*	Previous classification†	Response polarity	Striking response to small spots (100–200 μ m)	Spiking response to larger spots (600–1000 μ m)	Morphology
1 (22)	OFF Brisk Linear 1	OFF	Sustained‡	Sustained	Monostratified
2 (7)	OFF Brisk Linear 2	OFF	Sustained	Sustained	Monostratified, dye-coupled to amacrine cells
3 (36)	OFF Local Edge Detector	ON OFF for 50–200 μ m OFF for >200 μ m spots	Sustained	No response	Monostratified§
4 (25)	OFF Brisk Transient	OFF	Transient	Transient	Monostratified
5 (16)	ON Brisk Transient Nonlinear	ON	Transient	Transient	Monostratified
6 (20)	ON Brisk Transient Linear	ON	Transient	Very small response or no response	Monostratified
7 (28)	ON Brisk Sustained Linear	ON	Sustained	Transient	Diffusely stratified
8 (6)	ON Sluggish	ON	Sustained	Sustained	Monostratified
9 (10)	?	ON	Sustained	Sustained	Bi-stratified
10 (4)	ON Local Edge Detector	ON	Sustained	No response	?

* Numbers in brackets refer to the number of measured cells. (16 ON cells and 13 OFF cells were not classified.)

† The names refer to the classification of Amthor *et al.*²⁵

‡ A response was defined Transient if at the end of a 1-s flash there was no response. Otherwise it was defined as Sustained. ON cells were tested with white spots; OFF cells with black spots.

§ This cell type had a little wider ramification than the other monostratified cell classes. An example is shown in Fig. 4c.

(Fig. 3b). A function for postsynaptic inhibition is shown in cell type 3. The excitatory input has two components at light ON, a transient excitation inside the stimulus boundary and a more sustained excitation outside the stimulus boundary, along the edges of the stimulus. The wide and transient inhibitory input is coincident in time and space with the transient excitation and suppresses the excitation inside the stimulus boundary. This interaction results in a sharp representation of the outer boundary of the stimulus at the spiking pattern at light ON. In the other six cell types shown in Fig. 3 the spiking pattern was quite similar to the excitatory pattern that is brought to the IPL by means of the bipolar cells (Fig. 3b, compare columns 1 and 2). The inhibitory patterns that reflect the input from amacrine cells, however, hardly overlap in space and time with the excitatory patterns. This suggests that postsynaptic inhibition has only a small function in the formation of the diverse spatial representations for this specific stimulus.

One can consider the excitatory patterns measured in the gang-

lion cells as forming the initial property of each stratum, brought to the separate strata by bipolar cells with different space-time properties. The inhibitory patterns, which are markedly different from the excitatory patterns for each stratum shown in Fig. 3, suggest vertical interactions between strata. For example, many of the inhibitory patterns for the ON cells are driven by OFF bipolar cells²², and vice versa. In Fig. 5a the excitatory input to an ON ganglion cell type (cell type 9; $n = 6$) was completely blocked by 2-amino-4-phosphonobutyrate (APB)²³, a blocker of ON bipolar cell activity, whereas the postsynaptic inhibition was unaffected, suggesting that inhibition was driven by OFF bipolar cells (through amacrine cells). Notably, this ON ganglion cell type is bi-stratified, suggesting that dendritic processes ramifying in the OFF sublamina receive only inhibition, whereas dendritic processes ramifying in the ON sublamina receive only excitation. Conversely, cell type 1 was excited by the OFF system and inhibited by the ON system. This cell type is mono-stratified, so inhibition is probably mediated by amacrine processes

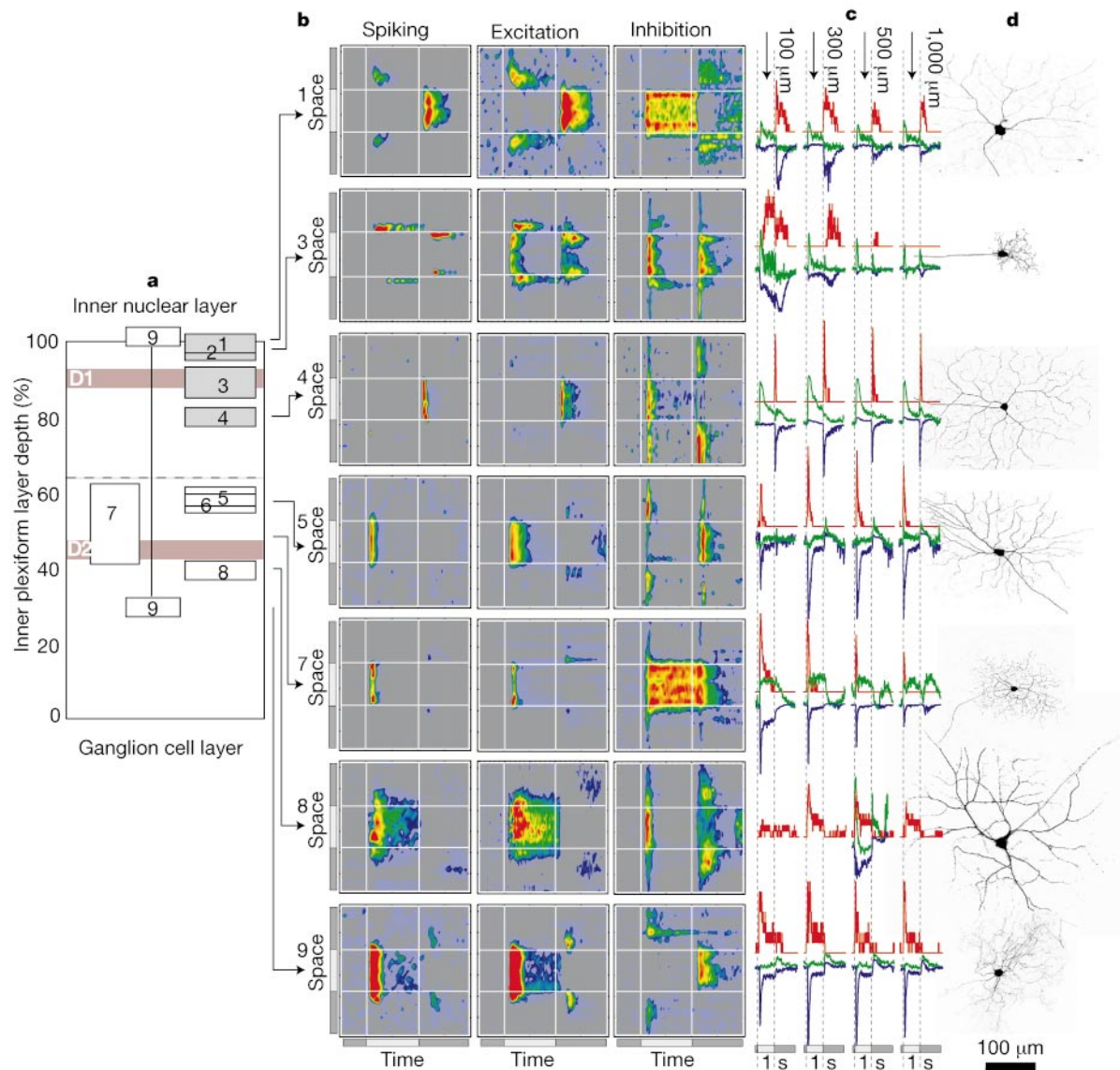


Figure 3 Seven of the ten parallel measured representations correlated with depth within the IPL. **a**, Identification of strata within the IPL. The depths of dendritic ramification for different ON (white) and OFF (grey) ganglion cells are shown. See Table 1 for names and classification of ganglion cells. D1 and D2 indicate the depth of dendritic branching of the directional selective ganglion cells. **b**, Spiking (first column), excitatory (second column) and inhibitory (third column) representations of the 600-μm flashed square. **c**, To help to relate these patterns to more conventional stimuli the spiking frequency (red, binned in

50-ms time windows), excitatory current (blue) and inhibitory current (green) responses of the same cells as in **b** to flashed white squares of different sizes (100, 300, 500, 1,000 μm, centred on the cell) are shown. Cell type 3 has a small spiking response to a 500-μm flashed square but a 600-μm flashed square centred on the cell completely abolishes the response, as shown in the centre row of the spiking pattern measurement. **d**, Morphology of measured ganglion cells projected to the plane of the ganglion layer. Scale bar, 100 μm.

that span several strata.

Other evidence for vertical communication between strata can be inferred by comparing the excitatory and inhibitory patterns with cell type 7 (Fig. 5b): excitation is transient and displays spatial-edge enhancement, whereas inhibition has a narrow sustained and a wide

transient component and does not display edge enhancement. This suggests that the bipolar cells that drive excitation are different from the ones that drive the inhibitory amacrine cells. Blocking ON activity in this cell type with APB eliminated excitation and also eliminated the sustained part of the inhibitory pattern within the

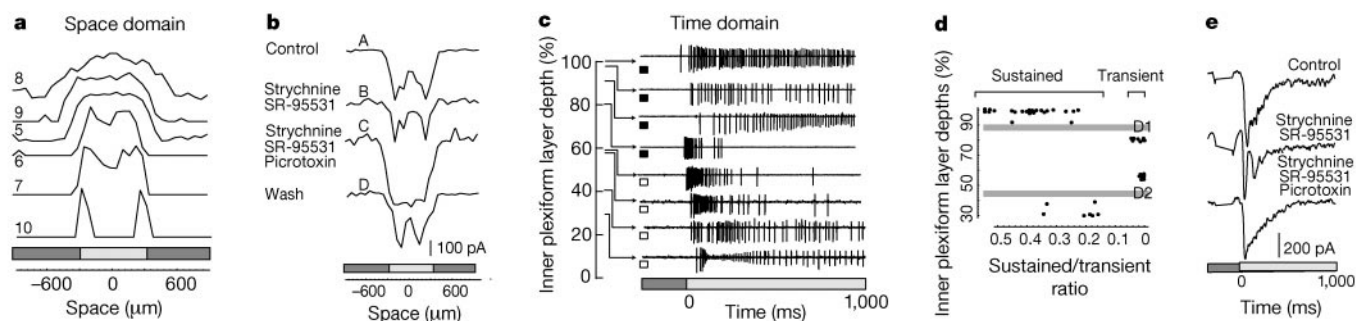


Figure 4 Order in the IPL. **a**, Range of spatial profiles of spike frequency for six different classes of ON ganglion cells measured at the peak of the response to a 600- μ m flashed square. **b**, Spatial profile of excitatory currents at the peak of the light response (600- μ m square stimulus) in cell type 7. Traces A and D, control conditions; trace B, glycine and GABA_A receptors blocked with 1–10 μ M strychnine and 5 μ M SR-95531; trace C, GABA_C receptors blocked additionally with 100 μ M picrotoxin. **c**, Individual responses from monostratified ganglion cells aligned according to the depth of their dendritic ramifications

within the IPL. The stimulus was a 100 μ m square: white for ON ganglion cells; black for OFF ganglion cells. **d**, The ratio of the peak spiking frequency and the frequency at the end of the 1-s flash is plotted against the depth of dendritic ramification for 48 monostratified ganglion cells. The stimulus was the same as in **c**. D1 and D2 are the same as in Fig. 3a. **e**, Excitatory current responses from a transient ganglion cell (cell type 5) under control conditions (upper trace), GABA_A and glycine receptor blockade (middle trace), and additional GABA_C blockade (lower trace). The stimulus was the same as in **c**.

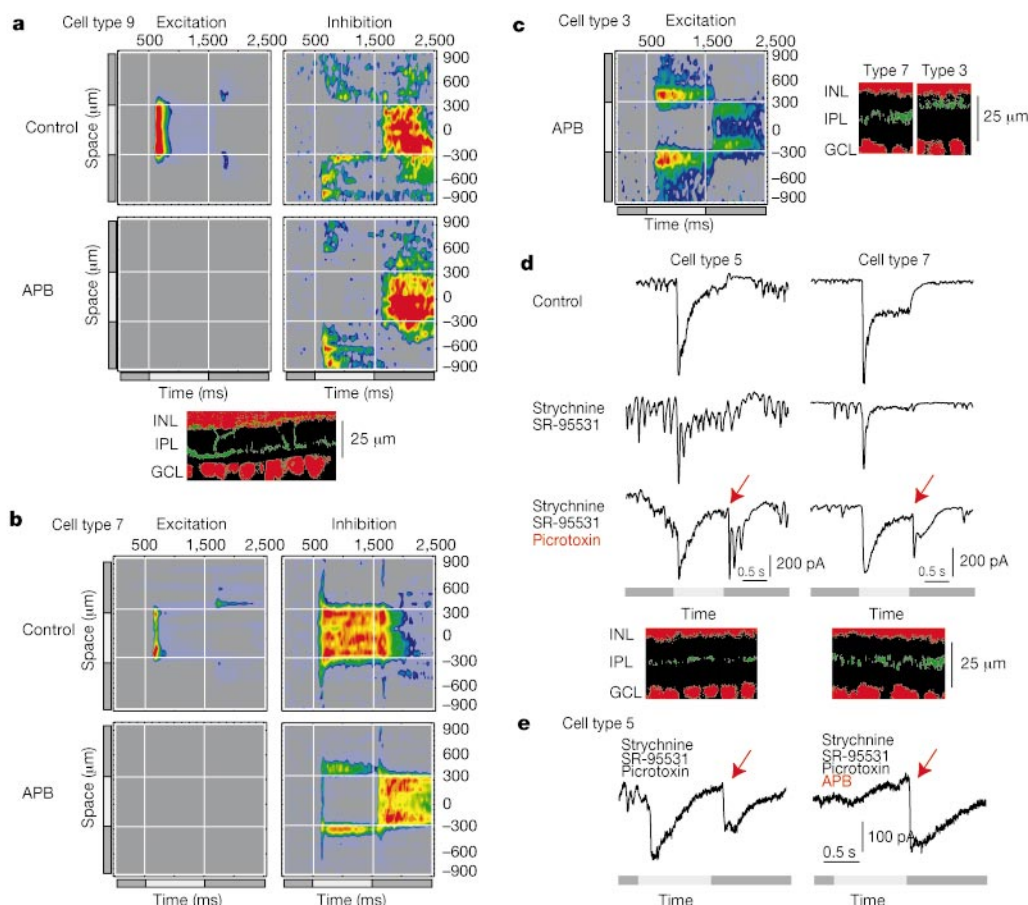


Figure 5 Vertical inhibition in the IPL. **a**, Excitatory and inhibitory input patterns to cell type 9 under control conditions (top plots) and in the presence of 20 μ M APB (bottom plots). The dendritic ramification of the measured cell is shown below the plots. **b**, Excitatory and inhibitory input patterns to cell type 7 under control conditions (top plots) and in the presence of 20 μ M APB (bottom plots). **c**, Left, an excitatory pattern in cell type 3 in the presence of 20 μ M APB. Right, the depth of dendritic ramifications is compared for the measured cells in **b** and **c**. **d**, Excitatory currents from cell type 5 (left) and cell type 7

(right). The stimulus was a 100- μ m light square. Upper trace, control conditions; middle trace, GABA_A and glycine receptor blockage with 10 μ M strychnine and 5 μ M SR-95531; lower trace, additional GABA_C receptor blockage with 100 μ M picrotoxin. Red arrows indicate the evoked OFF components. The depths of dendritic ramification for the measured cells are shown at the bottom. **e**, The OFF component (red arrows) evoked by inhibitory block (left trace) is APB resistant (right trace).

boundary of the flashed square at light ON. Two inhibitory components, however, remained intact: a wide transient component and a narrow sustained component outside the boundary of the flashed square at light ON. Comparing this narrow, sustained inhibitory component with the APB-resistant component of the excitatory pattern at the stratum of cell type 3 (Fig. 5c) suggests that inhibition is driven from this OFF stratum and conveyed by amacrine processes that span several strata.

Inhibitory interactions might also be involved in the separation of the functional properties of the strata. To test a function of feedback inhibition on the formation of the vertical stack of neural representations we monitored excitatory currents in two different ON ganglion cell types (cell types 5 and 7) with dendrites ramifying close to the ON–OFF border. In both cases the distal, neighbouring stratum was an OFF stratum. Figure 5d shows that under control conditions these cells receive excitation from bipolar cells only at light ON, but when feedback through GABA_C receptors was blocked, a sizeable OFF excitation was revealed (Fig. 5d: cell type 5, $n = 6$; cell type 7, $n = 8$). This OFF component was APB-resistant (Fig. 5e; $n = 5$), providing evidence that it was, in fact, driven by OFF bipolar cells. The unmasked OFF signal is either delivered by amacrine cells to ON bipolar cells through coupling or directly from OFF bipolar cells to ON ganglion cells. The direct coupling of the OFF signal from amacrine cells to the measured ganglion cell types is unlikely because dye coupling between cell types 5 and 7 and amacrine cells was not observed. These experiments indicate a structural and functional overlap of ON and OFF activity at the ON–OFF border, where the OFF component seems to be suppressed by inhibitory interactions mediated by feedback inhibition. Moreover, this experiment demonstrates that the separation of ON and OFF functional sublaminae is not rigid even after development²⁴ but is controlled by inhibition, leaving open the possibility for modulation of functional properties of ganglion cells.

We have shown that each stratum in the mammalian IPL embodies markedly different excitatory and inhibitory neural representations. In many cases, inhibition seems to be initiated from a different stratum than excitation and therefore must be carried vertically across the IPL. It is probable that diverse amacrine cell types perform this vertical association according to well-defined rules. It remains a challenge to determine these rules and to understand their functional significance.

Methods

Electrophysiology

We recorded from 193 ON or OFF ganglion cells in light-adapted whole-mount retinas of 2.5 kg New Zealand white rabbits using an Axopatch 200B amplifier (Axon Instruments). The loose, cell-attached electrode²⁵ was 3–4 M Ω , filled with Ames solution. The whole-cell electrode was 5–8 M Ω , filled with (in mM): 113 CsMeSO₄, 1 MgSO₄, 7.8×10^{-3} CaCl₂, 0.1 BABTA, 10 HEPES, 4 ATP-Na₂, 0.5 GTP-Na₃, 5 lidocaine *N*-ethyl bromide (QX314-Br), 7.5 neurobiotin chloride, pH 7.2. The retina was continuously perfused at 8–10 ml min⁻¹ with Ames (pH 7.4) solution at 36 °C, equilibrated with 95% O₂ and 5% CO₂. We determined spike frequency by counting spikes in a 50-ms time window, which moved in 1-ms time steps. We measured excitatory currents by clamping cells to E_{Cl} (the equilibrium potential for Cl⁻). The value of E_{Cl} was calculated to be –60 mV. E_{Cl} was also determined experimentally to be –65 to –60 mV by measuring the reversal potential for currents generated by inhibitory synaptic input, evoked by stimulating ganglion cells in the surrounds of their receptive fields.

Inhibitory currents were measured by clamping the membrane to 0 mV, which is the reversal potential for ionotropic glutamate receptors. This value was verified to lie between 0 and 10 mV by measuring the reversal potential for light-evoked currents when inhibitory inputs were blocked with strychnine and picrotoxin. By clamping the membrane to the actual measured values of the reversal potentials for excitation and inhibition we had reasonable assurance that we could adequately separate these two current components. This was further verified by measuring the spatial and temporal change of conductance and reversal potential from 25 ganglion cells clamped at six different holding potentials (between –80 and 20 mV). The measured excitatory and inhibitory currents are proportional to the change of conductance for the glutamate-gated and GABA- and glycine-gated channels, respectively. We note that the use of caesium-based intracellular solution and the internal blockage of voltage-gated sodium channels with QX314 was necessary to achieve correct voltage clamp. The data acquisition software (RED) was written by M. Wang. Data were analysed in Mathematica (Wolfram Research).

Confocal reconstruction

At the end of the whole-cell patch-clamp recordings we filled the measured ganglion cell with neurobiotin by supplying 1 nA depolarizing current steps (5 Hz) for 4 min. We then fixed and permeabilized the retina, followed by incubating it with streptavidin-Alexa Fluor 488 conjugate, ToPro3 and RNase. Optical sectioning of the whole-mount retina in the *z* axis at two different wavelengths (488 nm for ganglion cell morphology, 642 nm to image the nuclei in the ganglion cell layer and the inner nuclear layer) was performed on an MRC 1024 confocal microscope (Bio-Rad Laboratories). We analysed the stack of optical sections in Mathematica (Wolfram Research). The depth of dendritic ramification was determined relative to the borders of the nuclei of the ganglion cell layer and the inner nuclear layer. The depth of dendritic ramification was within 3–5% precision for a given ganglion cell type.

Light stimulus

An LCD panel, illuminated by a variable intensity (0–2,000 W) spatially homogenous lamp, projected the stimulus images onto the light-adapted whole-mount rabbit retina through an achromatic lens and the condenser. We focused the images on the photoreceptor layer. In all the experiments the background intensity was 4 nW m⁻². RED controlled the image parameters. The stimuli were 600- μ m (3.5 degrees of visual angle) flashed squares (+200% contrast) for the pattern measurements, 100, 300, 500 and 1,000- μ m flashed squares for the experiments in Fig. 3c, and 100- μ m flashed squares for the experiments in Figs 4c–e and 5d–e.

Received 26 October 2000; accepted 4 January 2001.

- Rodick, R. W. *The First Steps in Seeing* (Sinauer, Sunderland, Massachusetts, 1998).
- Albright, T. D., Jessell, T. M., Kandel, E. R. & Posner, M. I. Neural science: a century of progress and the mysteries that remain. *Neuron* **25**, S1–S55 (2000).
- Boycott, B. & Wässle, H. Parallel processing in the mammalian retina. The Proctor Lecture. *Invest. Ophthalmol. Vis. Sci.* **40**, 1313–1328 (1999).
- Euler, T. & Wässle, H. Immunocytochemical identification of cone bipolar cells in the rat retina. *J. Comp. Neurol.* **361**, 461–478 (1995).
- Euler, T., Schneider, H. & Wässle, H. Glutamate responses of bipolar cells in a slice preparation of the rat retina. *J. Neurosci.* **16**, 2934–2944 (1996).
- Cohen, E. & Sterling, P. Demonstration of cell types among cone bipolar neurons of cat retina. *Phil. Trans. R. Soc. Lond. B* **330**, 305–321 (1990).
- Maturana, H. R., Lettwin, J. Y., Pitts, W. H. & McCulloch, W. S. Physiology and anatomy of vision in the frog. *J. Gen. Physiol.* **43**, 129–175 (1960).
- Cajal, S. R. La rétine des vertébrés. *La Cellule* **9**, 119–257 (1893).
- Wu, S. M., Gao, F. & Maple, B. R. Functional architecture of synapses in the inner retina: segregation of visual signals by stratification of bipolar cell axon terminals. *J. Neurosci.* **20**, 4462–4470 (2000).
- Zhang, J., Li, W. & Massey, S. C. Cholinergic input to ganglion cells depends on depth in the IPL. *Invest. Ophthalmol.* **40**, S812 (1999).
- Famiglietti, E. V. Jr & Kolb, H. Structural basis for on- and off-center responses in retinal ganglion cells. *Science* **194**, 193–195 (1976).
- MacNeil, M. A. & Masland, R. H. Extreme diversity among amacrine cells: implications for function. *Neuron* **20**, 971–982 (1998).
- Roska, B., Nemeth, E., Orzo, L. & Werblin, F. S. Three levels of lateral inhibition: a space-time study of the retina of the tiger salamander. *J. Neurosci.* **20**, 1941–1951 (2000).
- Wässle, H. & Boycott, B. B. Functional architecture of the mammalian retina. *Phys. Rev.* **71**, 447–480 (1991).
- Fisher, S. K. & Boycott, B. B. Synaptic connections made by horizontal cells within the outer plexiform layer of the retina of the cat and the rabbit. *Proc. R. Soc. Lond. B* **186**, 317–331 (1974).
- Euler, T. & Wässle, H. Different contributions of GABA_A and GABA_C receptors to rod and cone bipolar cells in a rat retinal slice preparation. *J. Neurophys.* **79**, 1384–1395 (1998).
- Shields, C. R., Tran, M. N., Wong, R. O. & Lukasiewicz, P. D. Distinct ionotropic GABA receptors mediate presynaptic and postsynaptic inhibition in retinal bipolar cells. *J. Neurosci.* **20**, 2673–2682 (2000).
- Feigenspan, A., Wässle, H. & Bormann, J. Pharmacology of GABA receptor Cl⁻ channels in rat retinal bipolar cells. *Nature* **361**, 159–162 (1993).
- Qian, H. & Dowling, J. E. Novel GABA responses from rod-driven retinal horizontal cells. *Nature* **361**, 162–164 (1993).
- Enz, R., Brandstätter, J. H., Wässle, H. & Bormann, J. Immunocytochemical localization of the GABA_C receptor rho subunit in the mammalian retina. *J. Neurosci.* **16**, 4479–4490 (1996).
- Awatramani, G. B. & Slaughter, M. M. Origin of transient and sustained responses in ganglion cells of the retina. *J. Neurosci.* **20**, 7087–7095 (2000).
- Bloomfield, S. A. & Miller, R. F. A functional organization of ON and OFF pathways in the rabbit retina. *J. Neurosci.* **6**, 1–13 (1986).
- Slaughter, M. M. & Miller, R. F. 2-Amino-4-phosphonobutyric acid: a new pharmacological tool for retina research. *Science* **211**, 182–185 (1981).
- Wong, R. O. L. Retinal waves and visual system development. *Annu. Rev. Neurosci.* **22**, 29–47 (1999).
- Diamond, J. S. & Copenhagen, D. R. The relationship between light-evoked synaptic excitation and spiking behavior of salamander retinal ganglion cells. *J. Physiol.* **487**, 711–725 (1995).
- Amthor, F. R., Takahashi, E. S. & Oyster, C. W. Morphologies of rabbit retinal ganglion cells with concentric receptive fields. *J. Comp. Neurol.* **280**, 72–96 (1989).

Acknowledgements

We thank J. E. Dowling, M. M. Poo, R. H. Kramer and S. Picaud for their comments on the manuscript, and C. Chen for his assistance with the experiments. This study was supported by the ONR and NIH.

Correspondence and requests for materials should be addressed to F.W. (e-mail: werblin@socrates.berkeley.edu).

Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses

Rachel I. Wilson & Roger A. Nicoll

Departments of Cellular and Molecular Pharmacology and Physiology, University of California, San Francisco, California 94143, USA

Marijuana affects brain function primarily by activating the G-protein-coupled cannabinoid receptor-1 (CB1)¹⁻³, which is expressed throughout the brain at high levels⁴. Two endogenous lipids, anandamide and 2-arachidonylglycerol (2-AG), have been identified as CB1 ligands^{5,6}. Depolarized hippocampal neurons rapidly release both anandamide and 2-AG in a Ca²⁺-dependent manner⁶⁻⁸. In the hippocampus, CB1 is expressed mainly by GABA (γ -aminobutyric acid)-mediated inhibitory interneurons, where CB1 clusters on the axon terminal⁹⁻¹¹. A synthetic CB1 agonist depresses GABA release from hippocampal slices^{10,12}. These findings indicate that the function of endogenous cannabinoids released by depolarized hippocampal neurons might be to downregulate GABA release. Here we show that the transient suppression of GABA-mediated transmission that follows depolarization of hippocampal pyramidal neurons¹³ is mediated by retrograde signalling through release of endogenous cannabinoids. Signalling by the endocannabinoid system thus represents a mechanism by which neurons can communicate backwards across synapses to modulate their inputs.

Three properties of 'depolarization-induced suppression of inhibition' (DSI)¹³ suggested a role for cannabinoids in DSI. First, DSI, like endocannabinoid synthesis, requires Ca²⁺ influx into the postsynaptic neuron¹⁴. Second, DSI expression is probably presynaptic, as DSI does not affect the sensitivity of the postsynaptic membrane to iontophoresed GABA, or the quantal size of miniature GABA-mediated events^{13,15,16}, consistent with the localization of CB1 to GABA-containing axon terminals. Last, DSI is blocked by pertussis toxin¹⁷, implying a function for a G_i- or G_o-coupled receptor such as CB1. We therefore asked whether CB1 antagonists block DSI. We recorded from single CA1 pyramidal neurons in hippocampal slices in the whole-cell configuration. Figure 1a shows the transient depression of evoked inhibitory postsynaptic currents (eIPSCs) caused by a brief depolarizing step in the holding potential of a CA1 pyramidal neuron; incubating slices in the CB1 antagonist AM251 (Tocris) abolishes this effect. Figure 1b summarizes the effects of AM251 and another CB1 antagonist, SR141716 (National Institute on Drug Abuse), on DSI magnitude. SR141716 also blocked DSI induced by a train of action potentials (1 s, 20 Hz) rather than a voltage step (data not shown). Acute applications of SR141716 blocked the ability of a depolarizing step to depress eIPSC amplitude without affecting baseline eIPSCs (Fig. 1d). Furthermore, the effect of DSI on inhibitory transmission was mimicked by a synthetic CB1 agonist, WIN55212-2 (RBI), which acutely depressed baseline eIPSC amplitude (Fig. 1e). Pre-incubation in AM251 blocked this depression (eIPSC amplitude 95 $\pm$ 3% of baseline after 20 min of WIN55212-2 exposure (n = 6); data not shown, P < 0.05, compared with slices without AM251, paired t -test). WIN55212-2 did not affect the DSI-resistant component of the eIPSC (Fig. 1e), and thus occluded DSI.

The natural CB1 ligand 2-AG (Biomol) had only a small effect compared with the synthetic agonist (Fig. 2a), suggesting that 2-AG was not reaching sufficient concentrations in the slice. This is consistent with a recent report that both 2-AG and anandamide are substrates for an endogenous transporter that removes these ligands from the extracellular space^{18,19}. Application of AM404

(Tocris), an inhibitor of the anandamide/2-AG transporter, depressed baseline eIPSC amplitude without affecting the DSI-resistant component of the eIPSC (Fig. 2b). Pre-incubation of slices in SR141716 abolished the effects of AM404 (Fig. 2c). As AM404 is not itself a CB1 agonist¹⁸, these results imply that the endogenous substrate for the AM404-sensitive transporter is a cannabinoid that mimics and occludes DSI when allowed to accumulate in the slice. As a CB1 antagonist has no effect on baseline eIPSCs (Fig. 1d), tonic synthesis of endogenous cannabinoids must be normally balanced by uptake, which keeps extracellular cannabinoid levels below that required for CB1 activation. The kinetics of DSI were not changed as AM404 washed into the slice (Fig. 2d), suggesting that the AM404-sensitive transporter does not clear the retrograde signal rapidly enough to affect DSI decay. The kinetics of DSI probably reflect signal-transduction events inside the presynaptic terminal, or else passive diffusion of cannabinoids away from the site of release.

If an endogenous cannabinoid mediates DSI, then a CB1 agonist and DSI must depress GABA-mediated transmission by the same mechanism. We therefore asked whether DSI and WIN55212-2 act at a similar locus. In those experiments where WIN55212-2 depressed eIPSCs by at least 40%, the amplitude ratio of two closely

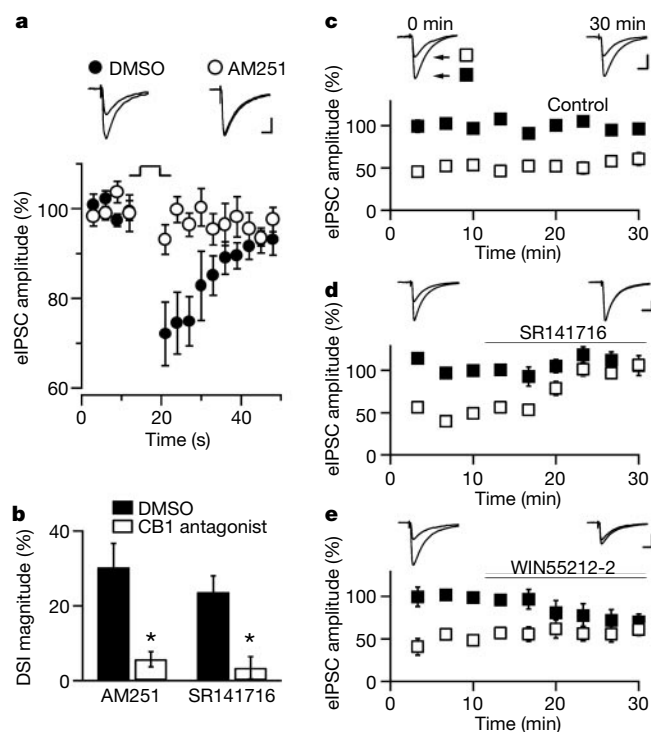


Figure 1 DSI requires endogenous cannabinoids. **a**, In pyramidal neurons (n = 14) from slices pre-incubated in AM251 (2 μ M), a 5-s depolarizing step results in little or no suppression of eIPSCs. Interleaved DMSO controls (n = 17) show a robust suppression. Average of 5 trials per cell. Insets show average IPSCs for the 10 s before and 10 s after the depolarizing step. **b**, Effect of AM251 (6 $\pm$ 2% versus 30 $\pm$ 6% average DSI in AM251 and DMSO alone, respectively) and SR141716 on DSI (2 μ M: 3 $\pm$ 3% versus 24 $\pm$ 4% average DSI in SR141716 (n = 11) and DMSO (n = 10) alone, respectively). Asterisk, P < 0.005. **c–e**, Control experiments (n = 7) (**c**) show stable DSI over 30 min, whereas SR141716 (n = 6) (**d**) blocks DSI by antagonism, and WIN55212-2 (800 nM: n = 5) (**e**) blocks DSI by occlusion (average DSI magnitude at 30 min in controls is 70 $\pm$ 10% of magnitude during baseline, compared with 11 $\pm$ 10% in SR141716 slices, and 19 $\pm$ 9% in WIN55212-2 slices; P < 0.005 for both treatments). Scale bars: **a**, **c–e**, 200 pA, 20 ms. The statistical test used was the paired t -test. Filled squares, eIPSC amplitude just before 5-s depolarization; open squares, eIPSC amplitude just after 5-s depolarization.

spaced eIPSCs (the paired-pulse ratio, PPR) was significantly increased (Fig. 3a). Similarly, DSI also reversibly increased PPR (Fig. 3b), which generally correlates with a local decrease in the probability of vesicular release from the axon terminal. In agreement with a previous study¹², we also found that WIN55212-2 decreases the frequency of Ca^{2+} -dependent miniature IPSCs (mIPSCs) recorded after blockade of action potentials by tetrodotoxin (TTX), and in the presence of high external potassium and Ca^{2+} concentrations (Fig. 3c, d). In the same conditions, but without TTX, depolarizing the postsynaptic cell induces a significant decrease in action-potential-driven spontaneous IPSCs (sIPSCs) (Fig. 3e). After adding TTX to the bath, depolarization elicits a similar depression of mIPSC frequency (Fig. 3e, f). Consistent with a presynaptic locus, DSI does not affect mIPSC amplitude ($100 \pm 3\%$ of baseline ($n = 5$); data not shown). Thus DSI, like the CB1 agonist, seems to act locally at the presynaptic terminal.

The conclusion that DSI is mediated by endogenous cannabinoids leads to three predictions about the properties of DSI. First, as both anandamide and 2-AG exit the cell by diffusion and/or passive transport, release of the retrograde signal in DSI should not require vesicular fusion⁸. We filled the postsynaptic cell by means of the recording electrode with botulinum toxin E light chain (BTE) while

monitoring DSI. BTE proteolytically cleaves both the synaptosome-associated proteins SNAP-25 and SNAP-23 (refs 20, 21); either of these is a necessary component of the minimal machinery required for all cellular membrane fusion²². We found that BTE does not affect the stability of DSI (Fig. 4a). As a control, we incubated recombinant SNAP-25 with BTE in our electrode-filling solution; the concentration of BTE used in our recordings was in ten-fold excess over that required to cleave all the SNAP-25 within 1 h (Fig. 4b). DSI is also unaffected by botulinum toxin B light chain (BTB) (Fig. 4a), which cleaves the vesicle-associated membrane protein VAMP2 and disrupts vesicular fusion in pyramidal cell dendrites when added to the solution in the recording electrode^{23,24}.

As the trigger for cannabinoid synthesis is cytoplasmic Ca^{2+} , a second prediction is that Ca^{2+} alone should be sufficient to trigger

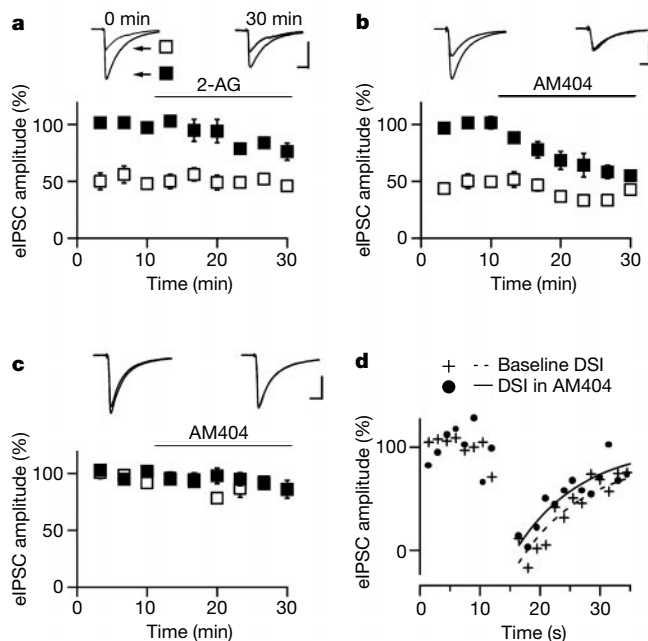


Figure 2 DSI is mimicked and occluded by blocking uptake of endogenous CB1 ligands. **a**, Effect of 2-AG on eIPSCs ($30 \mu\text{M}$: $76 \pm 7\%$ of baseline at 30 min, ($n = 6$) $P < 0.05$, compared with controls in Fig. 1c). **b**, AM404 significantly depresses eIPSCs ($20 \mu\text{M}$: $55 \pm 5\%$ of baseline at 30 min, ($n = 6$) $P < 0.001$, compared with controls in Fig. 1c) and partially occludes DSI (average DSI magnitude at 30 min is $39 \pm 11\%$ of magnitude during baseline, $P < 0.05$, compared with DSI from controls in Fig. 1c). **c**, Depression of eIPSCs by AM404 is antagonized by pre-incubation with SR141716 (eIPSC amplitude $85 \pm 4\%$ of baseline at 30 min, ($n = 5$) $P < 0.005$, compared to slices without SR141716). **d**, AM404 does not significantly affect the kinetics of DSI. Average time course of DSI during baseline period compared to the first 12 min of AM404 wash-in, during which DSI is being progressively occluded, is shown ($n = 6$, average of 4–5 baseline trials and 6 AM404 trials per cell). Lines are single exponential fits to the data (dashed line: baseline DSI, $\tau = 22.5$ s; solid line: DSI during first 12 min of AM404 wash-in, $\tau = 21.8$ s; not significantly different by a within-cell comparison of τ , $P > 0.7$, paired t -test). Scale bars: **a–c**: 200 pA, 20 ms. **d**: $\text{DSI} = \text{DSI}_{\text{max}} e^{-t/\tau}$.

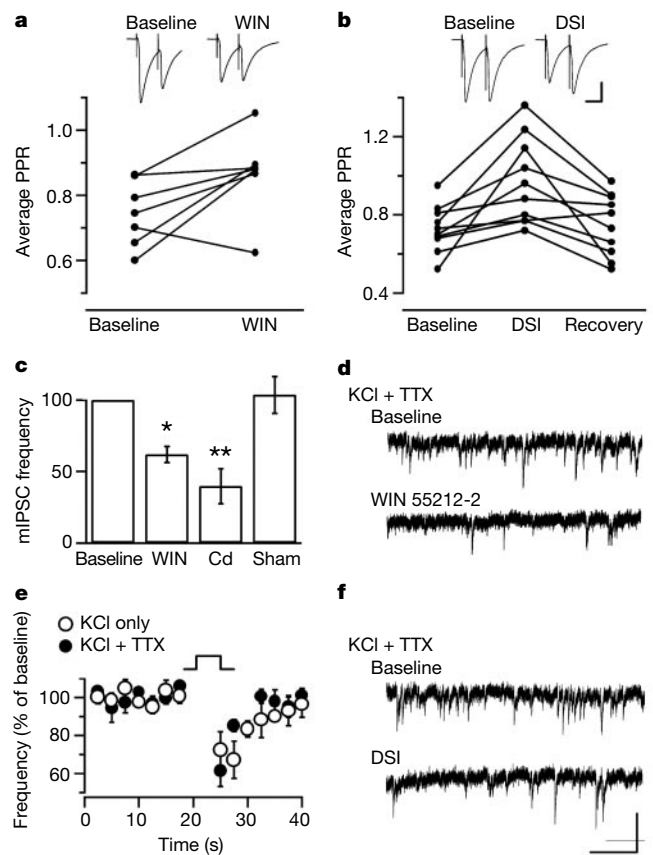


Figure 3 DSI and a CB1 agonist suppress IPSCs by the same mechanism. **a**, In experiments where WIN55212-2 depressed eIPSCs by at least 40%, the paired-pulse ratio (PPR) was significantly increased (inter-pulse interval 55 ms; ($n = 7$) $P < 0.05$). **b**, DSI reversibly increases the PPR of eIPSCs (inter-pulse interval 55 ms; ($n = 10$) $P < 0.01$). Scale bar: 200 pA, 40 ms. **c**, **d**, WIN55212-2 decreases the frequency of mIPSCs recorded in TTX, high external calcium, and high external potassium ($62 \pm 6\%$ of baseline frequency 20 min after wash-on; ($n = 6$) $P < 0.005$, compared with baseline or WIN55212-2; ($n = 5$) $P < 0.005$; double asterisk, $P < 0.05$). No change in mIPSC frequency occurred over 20 min when no drug was applied (Sham) ($104 \pm 13\%$ of baseline; ($n = 4$)). **e**, **f**, Under the same conditions of elevated calcium and potassium, but without TTX, depolarizing a pyramidal cell depresses action-potential-dependent sIPSCs (sIPSC frequency $74 \pm 7\%$ of baseline, $P < 0.001$; sIPSC amplitude $87 \pm 6\%$ of baseline, ($n = 5$) $P < 0.05$; data not shown). After adding TTX, depolarizing the same cell causes a similar decrease in mIPSC frequency ($77 \pm 4\%$ of baseline; ($n = 5$) $P < 0.005$). Scale bar: 50 pA, 1 s. The statistical test used was the paired t -test.

DSI. We find that the effect of a depolarizing step (Fig. 4c) is indistinguishable from the effect of uncaging Ca^{2+} from a photolabile chelator when the pyramidal neuron is not depolarized (Fig. 4d). Like DSI¹³, Ca^{2+} uncaging transiently depresses both the frequency and amplitude of sIPSCs (Fig. 4e). Ca^{2+} uncaging also depresses eIPSCs, an effect that is antagonized by AM251 (Fig. 4f). Like DSI, Ca^{2+} -induced depression is expressed presynaptically, as the amplitude of mIPSCs elicited by sucrose (150 mM) in the presence of TTX (1 μM) is unaffected ($97 \pm 4\%$ of baseline, measured 6 s after flash ($n = 8$); data not shown).

On the basis of reports that DSI is attenuated by the metabotropic glutamate receptor (mGluR) antagonist (S)- α -methyl-4-carboxyphenylglycine((S)-MCPG)²⁵, it has been suggested that the retro-

grade signal in DSI might be glutamate, although not all studies have found a role for mGluRs in DSI²⁶. Glutamate, or any other classical neurotransmitter, could be released from pyramidal cell somata by either vesicular fusion or by reversal of an electrogenic transporter on the plasma membrane. However, our data imply that neither of these processes can account for the release of the retrograde signal. We therefore re-examined the sensitivity of DSI to mGluR antagonists. The broad-spectrum, high-affinity mGluR antagonist LY341495 (Tocris) did not affect DSI (Fig. 4g). DSI was also resistant to (S)-MCPG (Tocris), although this dose significantly diminished the depression of eIPSCs by glutamate iontophoresed from a second electrode near the recorded cell (Fig. 4h). Our results are inconsistent with the hypothesis that glutamate, or indeed any classical neurotransmitter, is the primary retrograde signal in DSI. Although mGluRs seem not to be required for DSI, our results do not rule out the possibility that mGluRs might have a modulatory function in DSI. Indeed, recent data suggests that mGluR activation can enhance DSI in hippocampal slices (B. E. Alger, personal communication). Differences in experimental preparations could produce different levels of mGluR activation, and could explain the variability in the effect of

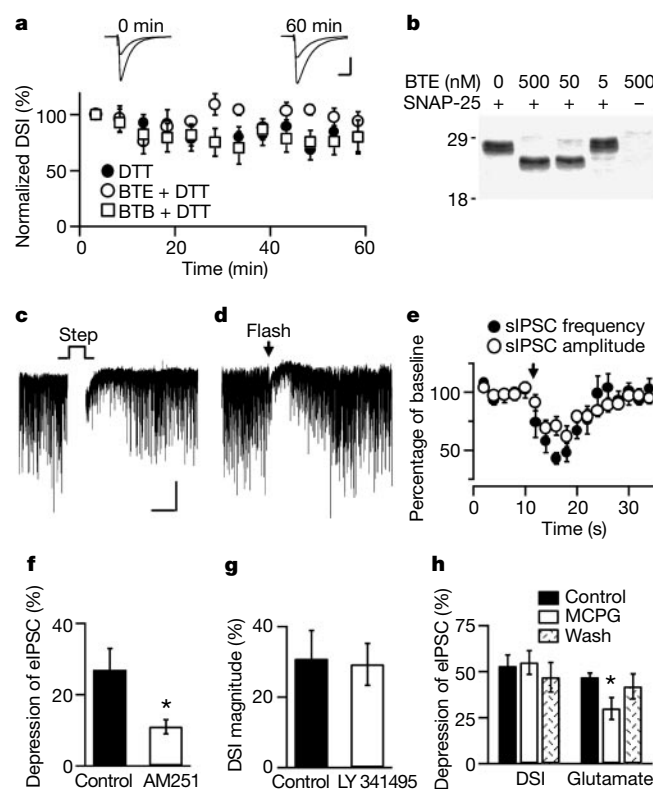


Figure 4 The postsynaptic properties of DSI are consistent with cannabinoids but inconsistent with a classical neurotransmitter. **a**, Filling pyramidal neurons with BTE ($n = 5$) or BTB ($n = 7$) (500 nM), plus DTT to maximize botulinum toxin activity, had no effect on the stability of DSI compared to DTT ($n = 8$) alone (DSI magnitude at 60 min was $94 \pm 9\%$ of the average magnitude during first 5 min for BTE and $80 \pm 15\%$ for BTB; versus $81 \pm 14\%$ for control). Scale bar: 200 pA, 20 ms. The mean absolute magnitudes of DSI were $52 \pm 3\%$ (BTE), $45 \pm 3\%$ (BTB) and $59 \pm 7\%$ (control). **b**, 50 nM BTE was sufficient to completely cleave in 60 min all the SNAP-25 diluted in a sample of electrode-filling solution. Relative molecular mass (K) shown on left. **c, d**, Depolarizing a pyramidal neuron (**c**) and photolysing caged calcium inside a pyramidal neuron (**d**) had similar effects on sIPSCs. Scale bar: 50 pA, 10 s. **e**, Both sIPSC frequency and sIPSC amplitude were transiently depressed by calcium uncaging (frequency $43 \pm 5\%$ of baseline, amplitude $71 \pm 8\%$ of baseline, both measured 6 s after flash; $n = 8$ cells). **f**, AM251 (2 μM) antagonizes the depression of eIPSCs elicited by calcium uncaging ($11 \pm 2\%$ average depression in AM251 ($n = 15$), versus $27 \pm 6\%$ average depression in DMSO alone, ($n = 16$)). Asterisk, $P < 0.01$, t -test. **g**, DSI is normal in slices pre-incubated and recorded in LY341495 (50 μM) ($29 \pm 6\%$ average DSI; $n = 10$) compared to controls ($31 \pm 8\%$ average DSI; $n = 10$). **h**, (S)-MCPG (5 mM) does not affect DSI ($53 \pm 6\%$ average DSI during baseline versus $55 \pm 6\%$ in (S)-MCPG; $n = 6$), although it reduces the depression of eIPSCs by iontophoresed glutamate in the same recordings ($47 \pm 2\%$ average depression during baseline versus $30 \pm 6\%$ in (S)-MCPG). Asterisk, $P < 0.05$, paired t -test.

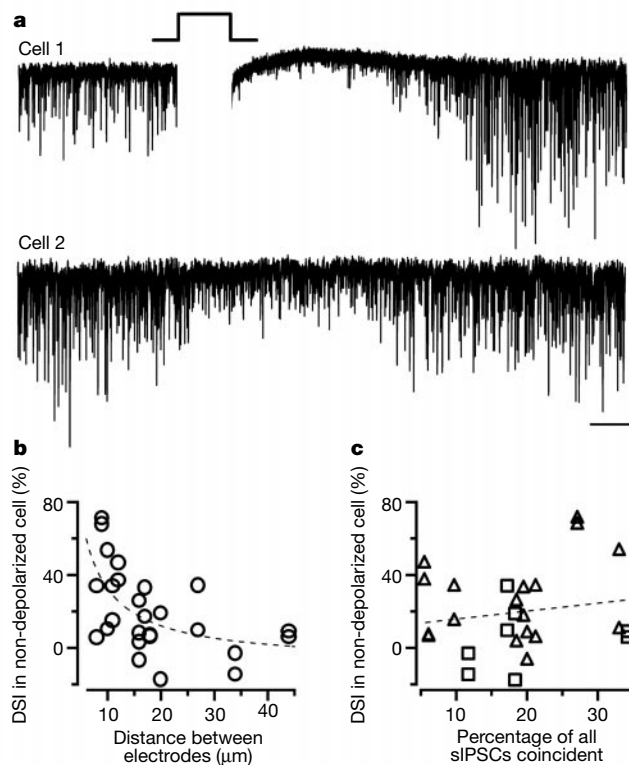


Figure 5 The retrograde signal in DSI can disinhibit nearby neurons. **a**, Depolarizing one CA1 pyramidal neuron for 5 s results in a transient suppression of sIPSCs (cell 1). The trace is blanked during the depolarizing step for clarity. A pyramidal cell 9 μm away (cell 2) also showed suppression of sIPSCs, beginning about 1 s after cell 1 was depolarized. Scale bar: 100 pA, 5 s. **b**, In 13 pyramidal cell pairs, the suppression of sIPSCs in the non-depolarized cell was steeply related to its distance from the depolarized cell, with little propagation beyond 20 μm ($r^2 = 0.34$, DSI in non-depolarized cell versus 1/distance; $P < 0.005$, analysis of variance). The dotted line is a linear fit of DSI in the non-depolarized cell versus 1/distance. **c**, In the same pairs, suppression of sIPSCs in the non-depolarized cell was not related to the degree of shared presynaptic input, assessed by the percentage of sIPSCs in both cells during the baseline period, which had a coincident event in the other cell ($r^2 = 0.03$, DSI in non-depolarized cell versus percentage sIPSCs coincident; not significant, analysis of variance). The dotted line is a linear fit of DSI in the non-depolarized cell versus percentage coincidence. Triangles, close pairs ($< 20 \mu\text{m}$ apart); squares, more distant pairs.

mGluR antagonists.

Endogenous cannabinoids are small diffusible molecules. A third prediction, therefore, is that DSI ought to spread. To test this, we recorded simultaneously in whole-cell mode from pairs of nearby pyramidal neurons. When depolarizing one of these two neurons depressed sIPSCs in that cell, a simultaneous depression of sIPSCs was often observed in the non-depolarized neighbour (Fig. 5a). We made 13 pairs of simultaneous recordings (26 neurons) where both neurons in the paired recording could produce DSI of their own IPSCs, signifying that these neurons were able to produce the retrograde signal and had presynaptic inputs sensitive to the signal. Of these 26 neurons, 10 showed an average suppression of inhibition of at least 20% immediately after the other neuron in the pair was depolarized. The degree of DSI propagation in these pairs is dependent on distance, with 9 of the 10 propagating pairs separated by 20 μm or less (Fig. 5b).

A DSI-like phenomenon in cerebellar Purkinje cells²⁷ and cultured myocytes²⁸ involves retrograde messengers that locally initiate a signalling cascade in the presynaptic axon terminal; this then propagates the depression through that axon to other synapses. If hippocampal DSI shared this property, then propagation should depend on the degree to which the two postsynaptic cells share some presynaptic inputs, as estimated by the percentage of sIPSCs in the two cells that are coincident. We find that the degree of propagation is uncorrelated with this variable (Fig. 5c), consistent with propagation occurring mainly through a diffusible molecule capable of spreading through a sphere of about 40 μm in diameter. This profile is consistent with the properties of the endocannabinoids, whose sphere of diffusion is limited by uptake, and hindered by their lipophilicity⁸.

Here we have placed cannabinoids in a small group of molecules identified as fast retrograde signals in the mature nervous system. Furthermore, our study represents the first identification of a physiological process mediated by endogenous brain cannabinoids. CB1-expressing interneurons are thought to control hippocampal oscillations in the theta and gamma frequency, and a CB1 agonist decreases the power of synchronous oscillations in hippocampal slices¹¹, suggesting that endogenous cannabinoids may modulate the synchronous spiking of hippocampal cells. Moreover, as decreasing GABA-mediated inhibition promotes long-term potentiation of glutamatergic synapses²⁹, DSI may facilitate learning under conditions of intense hippocampal recruitment. The fact that only strong depolarization can induce DSI means that relevant levels of cannabinoids probably occur in response to particularly intense stimuli. By contrast, exogenous cannabinoids such as marijuana might be predicted to tonically disinhibit the hippocampus, thereby destroying the information contained in the feedback loop of DSI and promoting a noisier, more random pattern of synaptic modification.

Note added in proof: It has come to our attention that two papers showing a role for endocannabinoids are being published in *Neuron* (Kreitzer, A. C. & Regehr, W. G. *Neuron*, **29**, 717–727; 2001 and Ohno-Shosaku, T., Maejima, T. & Kano, M. *Neuron*, **29**, 729–738; 2001). □

Methods

Slice preparation and electrophysiology

Transverse hippocampal slices (300 μm) were obtained from Sprague–Dawley rats (P16–30) and maintained in artificial cerebrospinal fluid (ACSF) for at least 1 h before recording. ACSF contained (in mM): 119 NaCl, 26 NaHCO₃, 10 glucose, 3 KCl, 2.5 CaCl₂, 2 MgSO₄, 1 NaH₂PO₄, 0.005 NBQX (Tocris), 0.002 CPP (Tocris), and was equilibrated with 95% O₂ and 5% CO₂ at 20–22 °C. We added carbachol (5–10 μM ; RBI) to the ACSF in some experiments (Figs 4c–e and 5) to increase sIPSC frequency¹⁷. Recording electrodes were filled with a solution of (in mM): 100 CsCH₃SO₃, 60 CsCl, 5 QX-314 chloride, 10 HEPES, 0.2 EGTA, 1 MgCl₂, 1 Mg-ATP, and 0.3 Na₃GTP (pH 7.3, 275 mOsm). Synaptic currents were filtered at 2 kHz and collected at 5 kHz. When series resistance exceeded 40 M Ω or input resistance fell below 100 M Ω , experiments were terminated. IPSCs were elicited using bipolar tungsten electrodes in or near CA1 stratum pyramidale. DSI tests, performed every 120 s, consisted of 30 stimuli at 0.33 Hz, with depolarization from

–60 mV to 0 mV for 5 s after the eighth stimulus. DSI was calculated using the mean of the three eIPSCs just before the depolarization ($\text{amp}_{\text{baseline}}$) and the three eIPSCs just after the depolarization (amp_{test}): $\text{DSI} (\%) = 100(1 - (\text{amp}_{\text{test}}/\text{amp}_{\text{baseline}}))$.

Pharmacology

Slices were pre-incubated with the drug AM251 (Figs 1a, b and 4f), SR141716 (Fig. 1b) or LY341495 (Fig. 4g) for 45–150 min, and recordings were performed in the presence of the same concentration. Drug-treated slices were interleaved with control slices from the same animal incubated for an equivalent period of time in the same concentration of the solvent (water, dilute NaOH, or DMSO) used to make the drug stock. Overall, about 60% of cells have significant DSI, with DSI in these cells averaging about 50% in magnitude; thus, DSI averaged about 30% in control experiments. For acute applications of drugs (Figs 1c–e and 2a–d), cells were discarded immediately if they did not show DSI of at least 30% during the 10-min baseline period.

Miniature IPSCs

In experiments examining mIPSCs (Fig. 3c–f) additional KCl (2.0–7.5 mM) was added to the ACSF at the beginning of the experiment until sIPSC frequency was at least 3 Hz. The ACSF in these experiments included high Ca²⁺ (7.5 mM total Ca²⁺) and low magnesium concentration (0.1 mM total Mg²⁺) to maximize the probability of release, whereas high total divalents served to decrease action-potential initiation. This permitted us, in DSI experiments (Fig. 3e–f), to record both action-potential-dependent sIPSCs and, after addition of TTX, mIPSCs at a frequency of 3–10 Hz without changing external KCl concentrations. Phosphate was omitted from the ACSF. In DSI experiments (Fig. 3e–f), cells showing DSI of eIPSCs measuring less than 30% on the first test (before addition of high external KCl) were discarded immediately. Although the cells selected for these experiments (Fig. 3e–f) all had robust DSI in normal ACSF (initial average DSI was $60 \pm 4\%$ before adding high potassium), elevating external potassium decreased the magnitude of DSI, such that the effect on sIPSC frequency and amplitude was modest, although highly significant. This is consistent with reports that depolarization decreases the efficacy of G-protein-mediated inhibition of voltage-dependent calcium currents³⁰, and may explain the failure of one study¹³ to find an effect of DSI on Ca²⁺-dependent mIPSCs, as even higher concentrations of external potassium were used.

A related issue is why we found an increase in the PPR during DSI, whereas other reports^{15,16} have not. In contrast to previous studies, we used a shorter inter-pulse interval (55 ms versus 200 ms) to maximize interaction between the two stimuli, and we used a paired *t*-test, which has better statistical resolution than an unpaired comparison. Finally, we note that under optimal conditions DSI might block completely transmission at the subset of terminals expressing CB1 and that, in these cases, a change in the PPR would not be expected.

Botulinum toxins and Ca²⁺ uncaging

BTE (500 nM; a gift from R. Scheller) light chain or BTB (500 nM; List Biolabs) light chain were dissolved in electrode-filling solution with 5 mM dithiothreitol (DTT). We performed recordings at 31 °C. Cells with DSI of less than 30% on the first test (1 min after break-in) were discarded immediately. For Fig. 4b, recombinant SNAP-25 (a gift from R. Scheller) was diluted to a concentration of 0.2 mg ml^{−1} in electrode-filling solution, along with 5 mM DTT and 0–500 nM BTE. After incubation for 60 min at 31 °C, proteins were separated on a 12% polyacrylamide gel and stained with Coomassie blue. In Ca²⁺-uncaging experiments (Fig. 4c–e) recording electrodes were filled with a solution containing (in mM): 100 CsCH₃SO₃, 78 tetraethylammonium chloride, 5 QX-314 chloride, 25 HEPES, 1 MgCl₂, 2 Mg-ATP, 0.3 Na₃GTP, 5 nitrophenyl-EGTA (Molecular Probes), 4 CsOH, 2 CaCl₂. Apamin (100 nM; CalBiochem) was included in the ACSF. Internal tetraethylammonium and external apamin were used to block Ca²⁺-dependent potassium conductances after uncaging. To uncage nitrophenyl-EGTA, cells were exposed for 800 ms to ultraviolet light from a 100 W mercury burner (Olympus) passed through a 25% neutral density filter.

Paired recordings

We made 39 pairs of simultaneous recordings from CA1 pyramidal neurons (78 neurons); 44 of the 78 (56%) were able to induce DSI of their own IPSCs. This percentage is consistent with the finding that only a subpopulation of interneurons expresses CB1 (ref. 10). Of these 44 neurons, 26 were members of a pair where both neurons were able to produce DSI of their own IPSCs. We analysed this set of recordings to determine under what conditions DSI might spread. The distance between neurons was defined as the distance between the tips of the two recording electrodes. Coincident sIPSCs were defined as events whose peak amplitude occurred within 7 ms of each other. A 7-ms window was chosen because longer windows did not substantially increase the frequency of calculated coincidence, whereas shorter windows did substantially decrease that frequency; also, average sIPSC rise time sometimes differed by 2–4 ms between the two cells, probably owing to different series resistances. As DSI of sIPSCs is a decrease in both frequency and amplitude of events, we express DSI here as a function of charge transfer normalized to time, that is, the sum of the areas of all sIPSCs in a period of time divided by the number of seconds in that period. Thus, expressed as a percentage decrease: $\text{DSI}(\%) = 100(1 - (\text{charge transfer}_{\text{test}}/\text{charge transfer}_{\text{baseline}}))$ where the baseline period was the 20 s immediately preceding the depolarization, and the test period was the 15 s immediately after the depolarization.

Received 11 December 2000; accepted 6 February 2001.

1. Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C. & Bonner, T. I. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346**, 561–564 (1990).
2. Caulfield, M. P. & Brown, D. A. Cannabinoid receptor agonists inhibit Ca current in NG108-15 neuroblastoma cells via a pertussis toxin-sensitive mechanism. *Br. J. Pharmacol.* **106**, 231–232 (1992).
3. Mackie, K. & Hille, B. Cannabinoids inhibit N-type calcium channels in neuroblastoma-glioma cells. *Proc. Natl Acad. Sci. USA* **89**, 3825–3829 (1992).
4. Herkenham, M. *et al.* Cannabinoid receptor localization in brain. *Proc. Natl Acad. Sci. USA* **87**, 1932–1936 (1990).
5. Devane, W. A. *et al.* Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* **258**, 1946–1949 (1992).
6. Stella, N., Schweitzer, P. & Piomelli, D. A second endogenous cannabinoid that modulates long-term potentiation. *Nature* **388**, 773–778 (1997).
7. Di Marzo, V. *et al.* Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* **372**, 686–691 (1994).
8. Di Marzo, V., Melck, D., Bisogno, T. & De Petrocellis, L. Endocannabinoids: endogenous cannabinoid receptor ligands with neuromodulatory action. *Trends Neurosci.* **21**, 521–528 (1998).
9. Tsou, K., Mackie, K., Sañudo-Peña, M. C. & Walker, J. M. Cannabinoid CB1 receptors are localized primarily on cholecystokinin-containing GABAergic interneurons in the rat hippocampal formation. *Neuroscience* **93**, 969–975 (1999).
10. Katona, I. *et al.* Presynaptically located CB1 cannabinoid receptors regulate GABA release from axon terminals of specific hippocampal interneurons. *J. Neurosci.* **19**, 4544–4558 (1999).
11. Hajos, N. *et al.* Cannabinoids inhibit hippocampal GABAergic transmission and network oscillations. *Eur. J. Neurosci.* **12**, 3239–3249 (2000).
12. Hoffman, A. F. & Lupica, C. R. Mechanisms of cannabinoid inhibition of GABA(A) synaptic transmission in the hippocampus. *J. Neurosci.* **20**, 2470–2479 (2000).
13. Pitler, T. A. & Alger, B. E. Postsynaptic spike firing reduces synaptic GABA responses in hippocampal pyramidal cells. *J. Neurosci.* **12**, 4122–4132 (1992).
14. Lenz, R. A., Wagner, J. J. & Alger, B. E. N- and L-type calcium channel involvement in depolarization-induced suppression of inhibition in rat hippocampal CA1 cells. *J. Physiol.* **512**, 61–73 (1998).
15. Alger, B. E. *et al.* Retrograde signalling in depolarization-induced suppression of inhibition in rat hippocampal CA1 cells. *J. Physiol.* **496**, 197–209 (1996).
16. Morishita, W. & Alger, B. E. Sr2+ supports depolarization-induced suppression of inhibition and provides new evidence for a presynaptic expression mechanism in rat hippocampal slices. *J. Physiol.* **505**, 307–317 (1997).
17. Pitler, T. A. & Alger, B. E. Depolarization-induced suppression of GABAergic inhibition in rat hippocampal pyramidal cells: G protein involvement in a presynaptic mechanism. *Neuron* **13**, 1447–1455 (1994).
18. Beltramo, M. *et al.* Functional role of high-affinity anandamide transport, as revealed by selective inhibition. *Science* **277**, 1094–1097 (1997).
19. Piomelli, D. *et al.* Structural determinants for recognition and translocation by the anandamide transporter. *Proc. Natl Acad. Sci. USA* **96**, 5802–5807 (1999).
20. Binz, T. *et al.* Proteolysis of SNAP-25 by types E and A botulinum neurotoxins. *J. Biol. Chem.* **269**, 1617–1620 (1994).
21. Leung, S. M., Chen, D., DasGupta, B. R., Whiteheart, S. W. & Apodaca, G. SNAP-23 requirement for transferrin recycling in Streptolysin-O-permeabilized Madin-Darby canine kidney cells. *J. Biol. Chem.* **273**, 17732–17741 (1998).
22. Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
23. Lledo, P. M., Zhang, X., Südhof, T. C., Malenka, R. C. & Nicoll, R. A. Postsynaptic membrane fusion and long-term potentiation. *Science* **279**, 399–403 (1998).
24. Lüscher, C. *et al.* Role of AMPA receptor cycling in synaptic transmission and plasticity. *Neuron* **24**, 649–658 (1999).
25. Morishita, W., Kirov, S. A. & Alger, B. E. Evidence for metabotropic glutamate receptor activation in the induction of depolarization-induced suppression of inhibition in hippocampal CA1. *J. Neurosci.* **18**, 4870–4882 (1998).
26. Morishita, W. & Alger, B. E. Differential effects of the group II mGluR agonist, DCG-IV, on depolarization-induced suppression of inhibition in hippocampal CA1 and CA3 neurons. *Hippocampus* **10**, 261–268 (2000).
27. Vincent, P. & Marty, A. Neighboring cerebellar Purkinje cells communicate via retrograde inhibition of common presynaptic interneurons. *Neuron* **11**, 885–893 (1993).
28. Cash, S., Zucker, R. S. & Poo, M. M. Spread of synaptic depression mediated by presynaptic cytoplasmic signaling. *Science* **272**, 998–1001 (1996).
29. Wigström, H. & Gustafsson, B. Facilitation of hippocampal long-lasting potentiation by GABA antagonists. *Acta Physiol. Scand.* **125**, 159–172 (1985).
30. Brody, D. L. & Yue, D. T. Relief of G-protein inhibition of calcium channels and short-term synaptic facilitation in cultured hippocampal neurons. *J. Neurosci.* **20**, 889–898 (2000).

Acknowledgements

We thank J. S. Isaacson and E. Schnell for suggesting a role for cannabinoids in DSI; R. S. Zucker for technical advice in the calcium uncaging experiments; and R. H. Scheller and Y. A. Chen for the gift of recombinant BTE light chain and SNAP-25. We are grateful for the comments on the manuscript contributed by D. S. Bredt, D. R. Copenhagen, R. H. Edwards, M. Frerking, D. Schmitz and M. P. Stryker. R.I.W. is supported by a National Science Foundation Graduate Research Fellowship. R.A.N. is a member of the Keck Center for Integrative Neuroscience and the Silvio Conte Center for Neuroscience Research. R.A.N. was supported by grants from the National Institutes of Health and the Bristol-Myers Squibb Corporation.

Correspondence and requests for materials should be addressed to R.A.N. (e-mail: nicoll@phy.ucsf.edu).

Ca²⁺ signalling between single L-type Ca²⁺ channels and ryanodine receptors in heart cells

Shi-Qiang Wang*, Long-Sheng Song*, Edward G. Lakatta* & Heping Cheng*†

* Laboratory of Cardiovascular Sciences, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224, USA

† National Laboratory of Biomembrane and Membrane Biotechnology, College of Life Sciences, Peking University, Beijing 100871, China

Ca²⁺-induced Ca²⁺ release is a general mechanism that most cells use to amplify Ca²⁺ signals^{1–5}. In heart cells, this mechanism is operated between voltage-gated L-type Ca²⁺ channels (LCCs) in the plasma membrane and Ca²⁺ release channels, commonly known as ryanodine receptors, in the sarcoplasmic reticulum^{3–5}. The Ca²⁺ influx through LCCs traverses a cleft of roughly 12 nm formed by the cell surface and the sarcoplasmic reticulum membrane, and activates adjacent ryanodine receptors to release Ca²⁺ in the form of Ca²⁺ sparks⁶. Here we determine the kinetics, fidelity and stoichiometry of coupling between LCCs and ryanodine receptors. We show that the local Ca²⁺ signal produced by a single opening of an LCC, named a 'Ca²⁺ sparklet', can trigger about 4–6 ryanodine receptors to generate a Ca²⁺ spark. The coupling between LCCs and ryanodine receptors is stochastic, as judged by the exponential distribution of the coupling latency. The fraction of sparklets that successfully triggers a spark is less than unity and declines in a use-dependent manner. This optical analysis of single-channel communication affords a powerful means for elucidating Ca²⁺-signalling mechanisms at the molecular level.

To visualize the trigger Ca²⁺ entry through single LCCs in intact rat ventricular myocytes, we used confocal imaging in the line-scan mode to detect the local Ca²⁺ signal directly beneath the voltage-clamped patch membrane⁷ (Fig. 1, inset). The sarcoplasmic reticulum (SR) Ca²⁺ release was paralysed by 10 mM caffeine and 10 μM thapsigargin, an SR Ca²⁺ pump blocker. When the pipette solution contained 20 mM Ca²⁺ and 10 μM FPL64176 (FPL), an LCC agonist that prolongs channel open time^{8,9}, patch depolarization from a holding potential of –50 mV to test potentials between –40 and 0 mV activated single-channel currents (*i*_{Ca}) in a voltage-dependent manner (Figs 1 and 2a). Each open event was accompanied by a discrete minuscule Ca²⁺ transient, which we named a 'Ca²⁺ sparklet', in close proximity to the patch membrane. Ca²⁺ sparklets were resistant to ryanodine (10 μM, 5 min, *n* = 3 patches), but were completely abolished by the LCC antagonist, nifedipine (2 μM, 3 min, *n* = 3 patches). These data indicate that Ca²⁺ sparklets originate from single LCC openings, and represent the first optical measurement of voltage-gated single-channel activity in intact cells.

Unlike Ca²⁺ sparks of ryanodine receptor (RyR) origin, which typically last for ~30 ms (ref. 6), Ca²⁺ sparklets have a variable duration: the onset and termination of Ca²⁺ sparklets correlated tightly with the open and closure of the channel, respectively (Fig. 1). To characterize sparklet properties in relation to Ca²⁺ influx, we measured the 'signal mass' of sparklets (that is, the space-time integral of local Δ*F*/*F*₀ of a sparklet in the line-scan image), and integrated the corresponding unitary Ca²⁺ influx (*q*_{Ca}). Over the voltage range from –30 mV to +10 mV, at which LCC openings can be clearly identified, there was a linear correlation between *q*_{Ca} (~8,000–100,000 Ca²⁺ ions) and sparklet signal mass (Fig. 2b). Hence, Ca²⁺ sparklets provide a faithful readout of LCC unitary current.

In heart cells, Ca²⁺ sparklets from single LCCs are expected to deliver trigger signal into the 12-nm junctional cleft^{10,11} to activate

the Ca^{2+} -gated RyRs and initiate Ca^{2+} sparks, as implied by ‘local control theory’^{12,13} and indirectly inferred by previous experiments^{14–17}. We therefore wanted to test directly whether a single LCC can trigger Ca^{2+} sparks, and if so, to determine the kinetics, fidelity and stoichiometry of the LCC–RyR coupling.

When we applied the G Ω -seal patch clamp and confocal imaging to myocytes with intact SR function (that is, in the absence of the SR inhibitors), we detected four Ca^{2+} sparks that were triggered unequivocally by i_{Ca} in 2 out of 28 patches that displayed LCC activities. Figure 3 shows that a large tail i_{Ca} (0.82 pA, 18.7 ms) on hyperpolarization from +30 mV to –110 mV evokes a typical Ca^{2+} spark with a latency of 4.2 ms. Notably, the LCC open duration in the presence of FPL outlasts the spark rise time (7 ms), suggesting that once triggered a Ca^{2+} spark evolves and terminates autonomously, independent of the duration of trigger i_{Ca} (ref. 14).

As LCC openings did not trigger a Ca^{2+} spark in most G Ω -seal patches, we thought that membrane deformation associated with the G Ω -seal might result in LCC–RyR uncoupling on most occasions. To preserve the functional integrity of the coupling apparatus, we implemented a ‘loose’ patch-clamp technique; that is, a low-resistance (20–50 M Ω) seal formed by gently pressing the patch pipette (5–7 M Ω) against the surface membrane without suction (Fig. 4a, inset). Although resolution of subpicoampere i_{Ca} is beyond the theoretical limits set by the inverse relation between seal resistance and electrical noise⁷ in such loose patches, detection of Ca^{2+} sparklets would nonetheless enable us to monitor single LCC activity by optical means.

In the presence of FPL, patch depolarization to ~0 mV evoked both high- and low-amplitude local Ca^{2+} events (Fig. 4a), giving rise to a bimodal amplitude distribution, with peaks around 20 and 100 nM, respectively (Fig. 4d, top panel). The low-amplitude events were resistant to 10 μM ryanodine, but were completely abolished by 2 μM nifedipine (Fig. 4a), and thus represent Ca^{2+} sparklets. When we decreased the pipette Ca^{2+} to 1 mM and omitted FPL, we could not detect a Ca^{2+} sparklet owing to a markedly reduced q_{Ca} . However, the high-amplitude events were still activated robustly (Fig. 4c), and exhibited a broad bell-shaped amplitude distribution similar to the ryanodine-sensitive population activated with 20 mM pipette Ca^{2+} (Fig. 4d). Hence, the high-amplitude events are Ca^{2+}

sparks of RyR origin.

The simultaneous visualization of sparklets and triggered sparks provides a unique opportunity to examine the control of SR Ca^{2+} release at the molecular level. Either inhibition of LCC influx by 2 μM nifedipine (Fig. 4b) or substitution of the pipette Ca^{2+} with 20 mM Ba^{2+} (Fig. 4c) extinguished spark activation, indicating that Ca^{2+} sparks are triggered by Ca^{2+} entry through LCCs, but not by voltage per se. To define the exact temporal relationship between the sparklet and the triggered spark, we obtained confocal images at an improved time resolution (0.7 ms per line) and a high signal-to-noise ratio (Fig. 5). Figure 5 shows that a slow-rising ‘foot’, which was not seen in spontaneous sparks, preceded the rapid upstroke of most triggered sparks. The similarity between the initial rate of rise of the feet and that of solitary sparklets (Fig. 6a, c) confirmed that the foot is a sparklet that triggers a spark. The latency (L) from the onset of a sparklet foot to its triggered Ca^{2+} spark was well described by a single-exponential function, with a time constant $\tau = 6.7$ ms (Fig. 6b). The exponentially fitted curve in Fig. 6b fully accounts for the triggered sparks with a sparklet foot ≤ 2.8 ms (counted in the

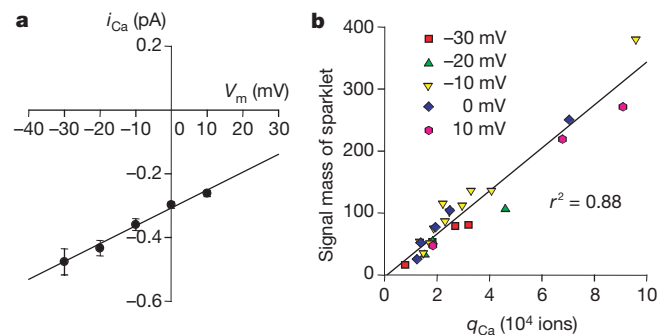


Figure 2 Quantitative relation between i_{Ca} and sparklets. **a**, The amplitude of i_{Ca} as a function of voltage exhibits a slope conductance of 5.11 ± 0.54 pS ($n = 6$ patches). **b**, There is a linear relationship ($r^2 = 0.88$) between ‘signal mass’ of sparklets ($\int \Delta F/F_0 dx dt$, in arbitrary units) and the corresponding unitary Ca^{2+} influx, q_{Ca} ($\int i_{\text{Ca}}/(2e) dt$).

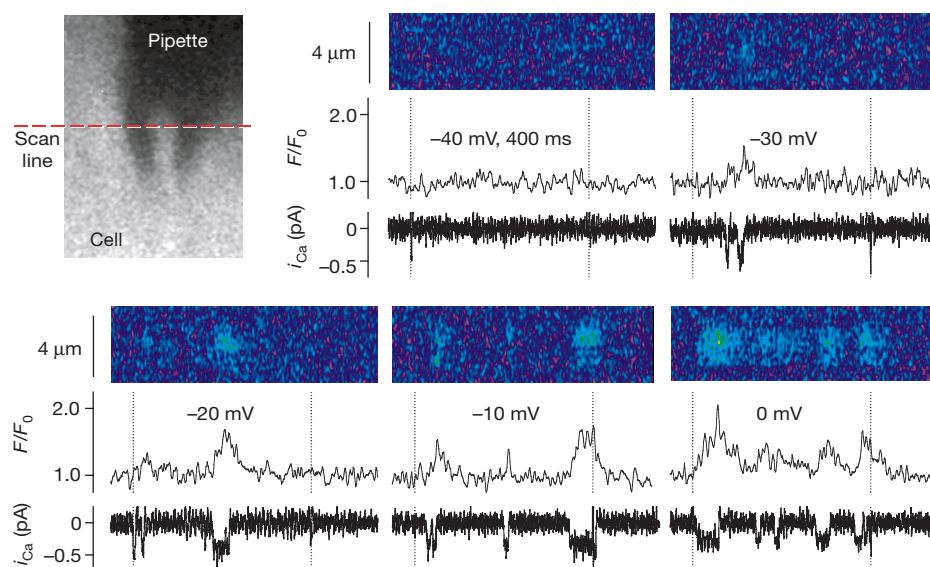


Figure 1 Confocal visualization of single LCC Ca^{2+} transients, ‘ Ca^{2+} sparklets’. Top, simultaneous recordings of confocal line-scan image; middle, normalized fluo-4 fluorescence; bottom, single LCC currents (i_{Ca}). Inset, experimental setting. The patch membrane was held at –50 mV and depolarized for 400 ms to voltages between –40 mV and 0 mV at 0.1 Hz. The membrane potential outside the patch was zeroed by a

high K^+ /0 Ca^{2+} solution. Note that LCC openings are associated with Ca^{2+} sparklets in a 1:1 fashion. These sparklets are apparently brighter (higher $\Delta F/F_0$) than those shown in Figs 4 and 5, mainly because of a low resting $[\text{Ca}^{2+}]$, and thus a low F_0 , in the absence of extracellular Ca^{2+} .

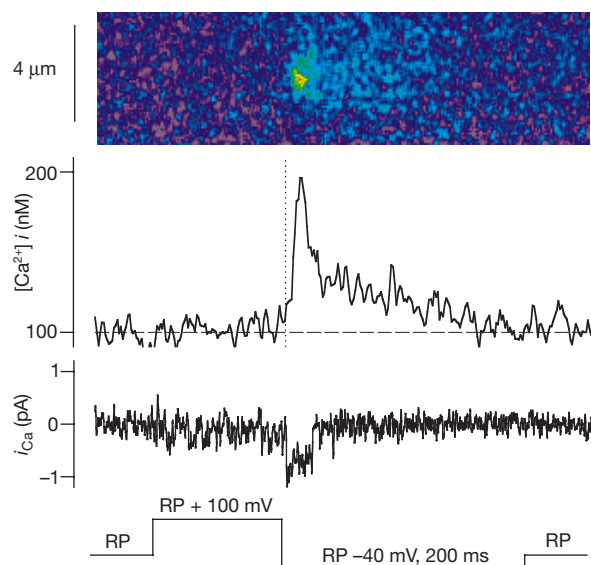


Figure 3 A Ca^{2+} spark triggered by single LCC opening. The cell was bathed in normal physiological saline (see Methods). The resting potential (RP) in rat ventricular myocytes is around -70 mV (ref. 28). A Ca^{2+} spark is evoked by a tail i_{Ca} on hyperpolarization from RP +100 mV to RP -40 mV. Note that the spark declines before L-type-channel current switches off.

first bin). The exponential distribution of coupling latency indicates that Ca^{2+} sparklets ignite Ca^{2+} sparks in a stochastic manner.

Because of stochastic variations of both LCC open time and LCC–RyR coupling latency, not every Ca^{2+} sparklet would be expected to trigger a Ca^{2+} spark. We therefore determined the LCC–RyR coupling fidelity (δ); that is, the fraction of sparklets that successfully trigger sparks. As shown in Fig. 6d, 71% of the first sparklets on depolarization triggered Ca^{2+} sparks, whereas 29% of them did not (that is, $\delta_1 = 0.71$), indicating a probabilistic rather than deterministic nature of LCC–RyR coupling. When the first sparklet did not trigger a spark, the conditional fidelity for the second sparklet remained unchanged ($\delta_{01} = 0.67$, $P > 0.05$ versus δ_1 , χ^2 -test). This suggests that, once terminated, a Ca^{2+} sparklet exerts no significant effect on subsequent LCC–RyR coupling, as the local $[\text{Ca}^{2+}]$ gradient dissipates in a fraction of a millisecond¹⁸. When the first sparklet did trigger a spark, however, the subsequent coupling fidelity was significantly reduced ($\delta_{11} = 0.30$, $P < 0.01$ versus δ_1 or δ_{01} , χ^2 -test), suggesting a use-dependent inactivation^{5,9} or ‘adaptation’¹⁹ of RyRs in intact cells.

A fundamental and unsettled issue regarding LCC–RyR coupling is the number of RyRs that are activated by a single LCC. In fact, whether a spark reflects a single RyR acting solo or a cluster of RyRs acting in concert remains a controversial issue^{6,20–25}. Because sparklets and triggered sparks share common Ca^{2+} buffering and diffusion microenvironments, Ca^{2+} sparklets of known i_{Ca} would offer a natural standard for calibration of the Ca^{2+} flux underlying a spark

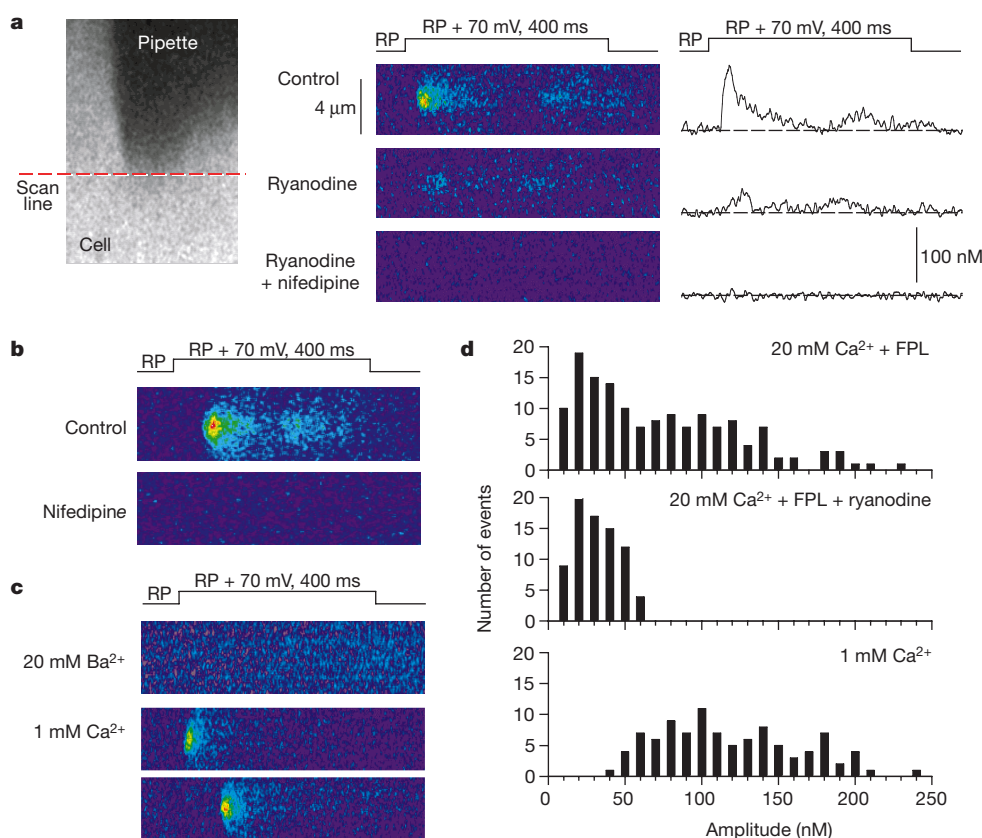


Figure 4 Sparklet–spark coupling under loose patch-clamp conditions. Inset, experimental setting. **a**, Top, both low-amplitude sparklets and high-amplitude sparks are evoked during patch depolarization from RP to ~ 0 mV. Middle, $10 \mu\text{M}$ ryanodine blocks Ca^{2+} sparks within ~ 3 – 5 min without affecting Ca^{2+} sparklet activity. Bottom, a subsequent application of $2 \mu\text{M}$ nifedipine, an LCC antagonist abolishes the remaining Ca^{2+} sparklets (bottom). **b**, Nifedipine alone ($2 \mu\text{M}$, ~ 3 – 5 min) abolishes both Ca^{2+} sparklet and the spark. **c**, Patch depolarization with Ba^{2+} as the charge carrier fails to evoke Ca^{2+} sparks; however, patch depolarization with 1 mM Ca^{2+} in the pipette in the

absence of FPL robustly evokes Ca^{2+} sparks. **d**, Amplitude histograms of local Ca^{2+} events under various experimental conditions. Top, intact SR with $10 \mu\text{M}$ FPL and 20 mM pipette Ca^{2+} ; middle, after ryanodine treatment; bottom, intact SR with 1 mM pipette Ca^{2+} without FPL. The broad modal distribution of spark amplitudes differs from the monotonic decaying distribution reported previously²², because the present experimental settings eliminate the ‘out-of-focus’ events that have confounded previous confocal study of sparks^{6,21,22}.

(j_{spark}). By comparing the initial rising rate of the triggered sparks (ν_2 , Fig. 6a), the sparklet foot (ν_1) or the solitary sparklet (ν_0), we found that local $[\text{Ca}^{2+}]$ rises about seven times faster during sparks than sparklets ($\nu_2/\nu_0 = 7.5$, $\nu_2/\nu_1 = 6.7$; Fig. 6c). Given an i_{Ca} amplitude of 0.30 pA at 0 mV, and assuming a linear relationship between the injecting Ca^{2+} flux and the rising rate of local $[\text{Ca}^{2+}]$, we estimated the j_{spark} to be ~ 2.1 pA. This value falls within the lower bound of model-derived estimation (2–10 pA)^{6,23,26}, but is well above the unitary Ca^{2+} current of cardiac RyRs (~ 0.35 – 0.6 pA measured in planar lipid bilayers under quasi-physiological ionic conditions)²⁷. Thus, our data suggest that an LCC typically triggers

about 4–6 RyRs to generate a Ca^{2+} spark.

In summary, we have developed the optical measurement of a single Ca^{2+} channel, and investigated the Ca^{2+} signalling between LCCs and RyRs at single-channel level. We have shown that a Ca^{2+} sparklet from a single LCC triggers SR Ca^{2+} release from four to six RyRs, directly validating the notion that cardiac SR Ca^{2+} release is under exquisite local control by single LCCs. Although Ca^{2+} sparklets tightly follow the open and close of LCCs, Ca^{2+} spark duration is relatively independent of the triggering sparklet. The LCC–RyR coupling exhibits a stochastic kinetics; the coupling fidelity is less than unity and use dependent. These findings provide insights into

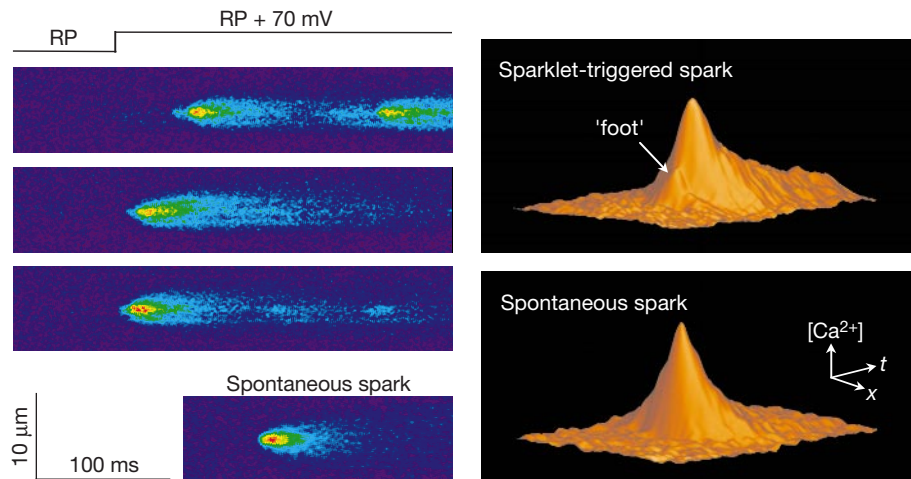


Figure 5 Visualization of the Ca^{2+} sparklets that immediately precede triggered Ca^{2+} sparks. Left, confocal images taken at a high temporal resolution (0.7 ms per line) and signal-to-noise ratio afforded by a LSM510 confocal microscope. The arrow indicates a

sparklet foot that triggered a spark. A representative spontaneous spark is shown at the bottom. Right, surface plot illustration of a triggered (top) and a spontaneous spark (bottom).

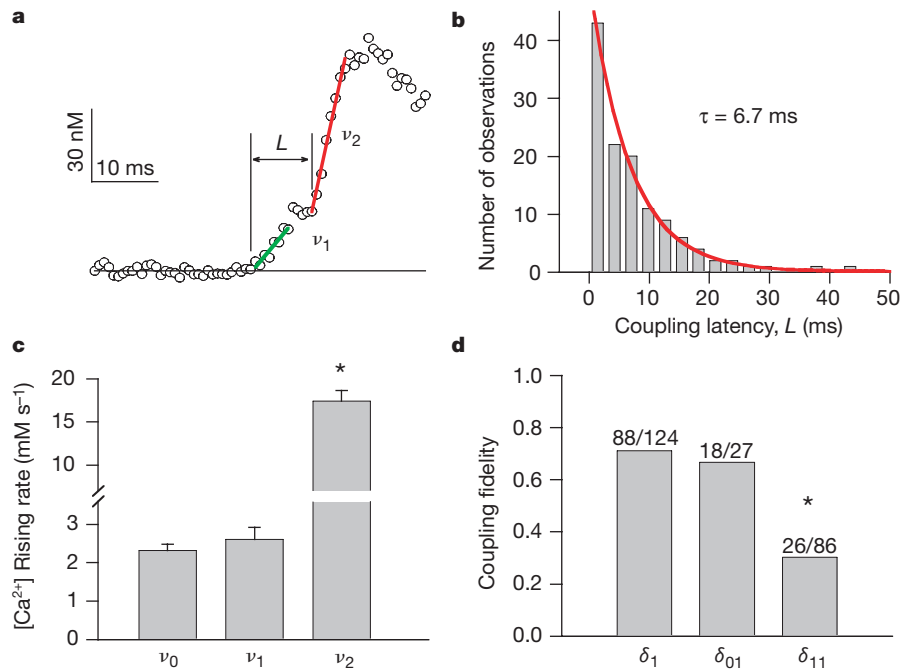


Figure 6 Latency and fidelity of LCC–RyR coupling. **a**, Illustration showing the measurement of the initial rising rate of the sparklet foot (ν_1) and triggered spark (ν_2), and their coupling latency (L). ν_1 and ν_2 were determined by linear fit of the first eight data points (for events with $L > 5.6$ ms). **b**, Histogram of the sparklet–spark coupling latency. Solid curve represents the least-squares single-exponential fit to the data. Time constant $\tau = 6.7$ ms. **c**, Rising rates for the solitary sparklets (ν_0 , $n = 14$), the foot (ν_1 , $n = 19$) and

the upstroke (ν_2 , $n = 19$) of triggered sparks. Note that $\nu_0 \approx \nu_1 \ll \nu_2$. Asterisk, $P < 0.01$ versus ν_0 or ν_1 . **d**, Fidelity of LCC–RyR coupling for the first events on depolarization (δ_1), the second events when the first sparklet did not trigger a spark (δ_{01}), and all events after a triggered spark (δ_{11}). Note that $\delta_1 \approx \delta_{01} > \delta_{11}$. Asterisk, $P < 0.01$ versus δ_1 or δ_{01} , χ^2 -test.

the molecular mechanisms of cardiac excitation–contraction coupling, and may have important implications for studying Ca^{2+} signalling in general. □

Methods

Confocal Ca^{2+} imaging

Enzymatically isolated adult rat ventricular myocytes were loaded with the Ca^{2+} indicator fluo-4-AM as described^{6,28}. Confocal line-scan imaging was performed using a Zeiss LSM 410 or LSM510 confocal microscope equipped with an argon laser (488 nm) and a $\times 40$, 1.3NA oil immersion objective. Line-scan images were acquired at sampling rates of 0.7 or 1.4 ms per line and 0.07 μm per pixel, with radial and axial resolutions of 0.4 and 1.0 μm , respectively. Digital image analysis was performed using IDL software (Research Systems) and customer-devised routines^{6,22}. In SR-paralysed cells, local Ca^{2+} transients were indexed by the normalized local fluorescence, F/F_0 (F_0 refers to the level before depolarization). In cells with intact SR function, local Ca^{2+} transients were determined using the following formula⁶:

$$[\text{Ca}^{2+}] = k_d R / (k_d / [\text{Ca}^{2+}]_{\text{rest}} + 1 - R)$$

where $R = F/F_0$, the resting Ca^{2+} concentration $[\text{Ca}^{2+}]_{\text{rest}} = 100 \text{ nM}$, and the dissociation constant $k_d = 1.1 \mu\text{M}$ (in accordance with the value for fluo-3 in cells²⁹).

Cell-attached patch clamp

Cell-attached patch clamps were established in either G Ω -seal or loose-seal configuration⁷, with glass patch pipettes of 5–7 M Ω . In G Ω -seal patches, unitary Ca^{2+} currents were recorded by using a cooled capacitor feedback headstage (CV203B) and an Axopatch 200 B amplifier (Axon Instruments). Current records were low-pass filtered at 1 kHz, digitized at 5 kHz, and are presented with leak and capacity currents eliminated by subtraction of smooth functions, as described⁹. For loose-seal patch clamp experiments, membrane potential was determined by proportionally dividing the test voltages between the pipette resistance (5–7 M Ω) and the seal resistance (20–50 M Ω). The surface-charge shielding effect by high Ca^{2+} or Ba^{2+} was not corrected.

Solutions

The standard patch pipette filling solution contained (in mM): 120 tetraethylammonium chloride, 20 CaCl_2 , 0.01 tetrodotoxin, 0.01 FPL64176 and 10 HEPES, pH 7.4, unless specified otherwise. In experiments shown in Fig. 1, the extracellular solution outside the patch membrane contained (in mM): 100 potassium aspartate, 40 KCl, 10 HEPES, 1 EGTA, 5 Mg^{2+} -ATP, 10 caffeine and 0.01 thapsigargin, pH 7.3. In other experiments, the standard superfusion solution contained (in mM): 135 NaCl, 1 CaCl_2 , 4 KCl, 1 MgCl_2 , 10 glucose, 10 HEPES, pH 7.4. All experiments were performed at room temperature (23–25 °C).

Statistics

Data were expressed as mean $\pm$ s.e.m. Student's *t*-test, paired *t*-test and χ^2 -test were applied when appropriate. A *P* value less than 0.05 was considered statistically significant.

Received 31 October 2000; accepted 10 January 2001.

- Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M. & Poo, M.-M. Calcium signalling in the guidance of nerve growth by netrin-1. *Nature* **403**, 93–98 (2000).
- Clapham, D. E. Calcium signaling. *Cell* **80**, 259–268 (1995).
- Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Bers, D. M. & Perez-Reyes, E. Ca channels in cardiac myocytes: structure and function in Ca influx and intracellular Ca release. *Cardiovasc. Res.* **42**, 339–360 (1999).
- Fabiato, A. Time and calcium dependence of activation and inactivation of calcium-induced release of calcium from the sarcoplasmic reticulum of a skinned canine cardiac Purkinje cell. *J. Gen. Physiol.* **85**, 247–289 (1985).
- Cheng, H., Lederer, W. J. & Cannell, M. B. Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science* **262**, 740–744 (1993).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
- Kunze, D. L. & Rampe, D. Characterization of the effects of a new Ca^{2+} channel activator, FPL 64176, in GH3 cells. *Mol. Pharmacol.* **42**, 666–670 (1992).
- Sham, J. S. et al. Termination of Ca^{2+} release by a local inactivation of ryanodine receptors in cardiac myocytes. *Proc. Natl Acad. Sci. USA* **95**, 15096–15101 (1998).
- Franzini-Armstrong, C., Protasi, F. & Ramesh, V. Shape, size, and distribution of Ca^{2+} release units and couplings in skeletal and cardiac muscles. *Biophys. J.* **77**, 1528–1539 (1999).
- Carl, S. L. et al. Immunolocalization of sarcolemmal dihydropyridine receptor and sarcoplasmic reticular triad and ryanodine receptor in rabbit ventricle and atrium. *J. Cell. Biol.* **129**, 672–682 (1995).
- Stern, M. D. Theory of excitation-contraction coupling in cardiac muscle. *Biophys. J.* **63**, 497–517 (1992).
- Niggli, E. & Lederer, W. J. Voltage-independent calcium release in heart muscle. *Science* **250**, 565–568 (1990).
- Cannell, M. B., Cheng, H. & Lederer, W. J. The control of calcium release in heart muscle. *Science* **268**, 1045–1049 (1995).
- López-López, J. R., Shacklelock, P. S., Balke, C. W. & Wier, W. G. Local calcium transients triggered by single L-type calcium channel currents in cardiac cells. *Science* **268**, 1042–1045 (1995).

- Shorofsky, S. R., Izu, L., Wier, W. G. & Balke, C. W. Ca^{2+} sparks triggered by patch depolarization in rat heart cells. *Circ. Res.* **82**, 424–429 (1998).
- Santana, L. F., Cheng, H., Gómez, A. M., Cannell, M. B. & Lederer, W. J. Relation between the sarcolemmal Ca^{2+} current and Ca^{2+} sparks and local control theories for cardiac excitation-contraction coupling. *Circ. Res.* **78**, 166–171 (1996).
- Soeller, C. & Cannell, M. B. Numerical simulation of local calcium movements during L-type calcium channel gating in the cardiac diad. *Biophys. J.* **73**, 97–111 (1997).
- Györke, S. & Fill, M. Ryanodine receptor adaptation: control mechanism of Ca^{2+} -induced Ca^{2+} release in heart. *Science* **260**, 807–809 (1993).
- Lipp, P. & Niggli, E. Fundamental calcium release events revealed by two-photon excitation photolysis of caged calcium in guinea-pig cardiac myocytes. *J. Physiol. (Lond.)* **508**, 801–809 (1998).
- Bridge, J. H., Ershler, P. R. & Cannell, M. B. Properties of Ca^{2+} sparks evoked by action potentials in mouse ventricular myocytes. *J. Physiol. (Lond.)* **518**, 469–478 (1999).
- Cheng, H. et al. Amplitude distribution of calcium sparks in confocal images: theory and studies with an automatic detection method. *Biophys. J.* **76**, 606–617 (1999).
- González, A. et al. Involvement of multiple intracellular release channels in calcium sparks of skeletal muscle. *Proc. Natl Acad. Sci. USA* **97**, 4380–4385 (2000).
- Schneider, M. F. Ca^{2+} sparks in frog skeletal muscle: generation by one, some, or many SR Ca^{2+} release channels? *J. Gen. Physiol.* **113**, 365–372 (1999).
- Bootman, M. D. & Berridge, M. J. The elemental principles of calcium signaling. *Cell* **83**, 675–678 (1995).
- Blatter, L. A., Hüser, J. & Ríos, E. Sarcoplasmic reticulum Ca^{2+} release flux underlying Ca^{2+} sparks in cardiac muscle. *Proc. Natl Acad. Sci. USA* **94**, 4176–4181 (1997).
- Mejia-Alvarez, R., Kettlun, C., Ríos, E., Stern, M. & Fill, M. Unitary Ca^{2+} current through cardiac ryanodine receptor channels under quasi-physiological ionic conditions. *J. Gen. Physiol.* **113**, 177–186 (1999).
- Xiao, R. P., Spurgeon, H. A., O'Connor, F. & Lakatta, E. G. Age-associated changes in beta-adrenergic modulation on rat cardiac excitation-contraction coupling. *J. Clin. Invest.* **94**, 2051–2059 (1994).
- Harkins, A. B., Kurebayashi, N. & Baylor, S. M. Resting myoplasmic free calcium in frog skeletal muscle fibers estimated with fluo-3. *Biophys. J.* **65**, 865–881 (1993).

Acknowledgements

We thank W. J. Lederer, M. B. Cannell, M. D. Stern, E. Ríos, J. S. K. Sham, S. J. Sollott, I. Josephson and R. P. Xiao for critical comments on the manuscript; H. A. Spurgeon for technical support; and A. Erauth for secretarial assistance. This work was supported by the NIH intramural research program (to E.G.L. and H.C.) and grants from the National Natural Science Foundation of China (H.C.).

Correspondence and requests for materials should be addressed to H.C. (e-mail: chengp@grc.nia.nih.gov).

Drought-induced guard cell signal transduction involves sphingosine-1-phosphate

Carl K.-Y. Ng^{*}, Kathryn Carr[†], Martin R. McAinsh^{*}, Brian Powell[†] & Alistair M. Hetherington^{*}

^{*} Department of Biological Sciences, Institute of Environmental and Natural Sciences, Lancaster University, Bailrigg, Lancaster LA1 4YQ, UK
[†] Avecia Limited, Hexagon House, Blackley, Manchester M9 8ZS, UK

Stomata form pores on leaf surfaces that regulate the uptake of CO_2 for photosynthesis and the loss of water vapour during transpiration¹. An increase in the cytosolic concentration of free calcium ions ($[\text{Ca}^{2+}]_{\text{cyt}}$) is a common intermediate in many of the pathways leading to either opening or closure of the stomatal pore^{2,3}. This observation has prompted investigations into how specificity is controlled in calcium-based signalling systems in plants. One possible explanation is that each stimulus generates a unique increase in $[\text{Ca}^{2+}]_{\text{cyt}}$, or 'calcium signature', that dictates the outcome of the final response⁴. It has been suggested that the key to generating a calcium signature, and hence to understanding how specificity is controlled, is the ability to access differentially the cellular machinery controlling calcium influx and release from internal stores^{2–5}. Here we report that sphingosine-1-phosphate is a new calcium-mobilizing molecule in plants. We show that after drought treatment sphingosine-1-phosphate levels increase, and we present evidence that this molecule is

involved in the signal-transduction pathway linking the perception of abscisic acid to reductions in guard cell turgor.

In animal cells, sphingosine-1-phosphate (S1P) has been reported to be involved in the mediation of numerous cellular responses including proliferation, survival, cytoskeletal organization, motility, differentiation and neurite retraction and rounding^{6–8}. As a first step to establishing whether S1P is an endogenous calcium-mobilizing compound in plants, we confirmed its presence in plant extracts. In animals, the predominant sphingoid base is sphingosine (*trans*-4-sphingenine, d18:1 (ref. 4)) with lesser amounts of sphinganine (d18:0) and 4-hydroxysphinganine (t18:0)⁹. In contrast, sphingoid bases from plants are predominantly sphinganine, 4-hydroxysphinganine and *cis* and *trans* isomers of 8-sphingenine (d18:1 (ref. 8)), 4,8-sphingadiene (d18:2 (refs 4, 8)) and 4-hydroxy-8-sphingenine (t18:1 (ref. 10)). (Note that d and t denote a dihydroxy and trihydroxy base, respectively.)

We extracted lipids from mature, fully expanded leaves of *Commelina communis*. The lipid extract was fractionated by high-performance liquid chromatography (LC) and the fraction corresponding to S1P was eluted. The putative S1P fraction was subjected to tandem mass spectrometry (MS/MS) in both positive and negative ionization modes and the corresponding product ion spectra were compared with those obtained using authentic S1P. The positive-ion LC/MS/MS product ion spectrum of *m/z* 380 (M+H)⁺ from the lipid extract (Fig. 1a) matched the spectrum obtained from authentic S1P (Fig. 1b). The negative-ion LC/MS/MS product ion spectrum of *m/z* 378 (M-H)⁻ from the lipid extract showed a single peak at *m/z* 79 corresponding to the phosphoryl group, matching that obtained from authentic S1P (data not shown). The results of this experiment confirmed the presence of S1P in plants. Notably, our results also challenge the current consensus that plants are devoid of sphingoid bases with the Δ 4-double bond¹⁰. We estimate that the levels of S1P in leaves range from 5 to 46 pg per g dry weight (on the basis of a recovery of ~16%). In addition, when plants were subject to drought treatments, we observed a 1.3–2.4-fold increase in S1P compared with levels in well-watered controls in four out of six pairwise comparisons (see Supplementary Information Table 1). Although the conditions for S1P extraction have yet to be fully optimized for plant tissue, these data suggest that decreased water availability might be associated with increased S1P levels in leaves.

We next tested the ability of S1P to modulate guard cell turgor by measuring changes in stomatal apertures after incubation of epidermal strips of *C. communis* in buffer containing S1P. S1P (4–6 μ M) promoted the closure of opened stomata (Fig. 2a). The effect

of S1P is reversible, as determined by wash-out experiments, indicating that the observed effect was not due to irreversible cytotoxicity. In addition, guard cells treated with S1P were morphologically normal and showed a normal pattern of organelle distribution, providing further evidence that the effect of S1P was not due to cytotoxicity (data not shown). Studies using animal cells showed that the presence of the Δ 4-double bond is essential for S1P to mediate its intracellular effects¹¹. To test whether the effect of S1P is specific and dependent on the Δ 4-double bond, we used the structurally similar sphingoid base, dihydro-S1P, which is derived from S1P by hydrogenation of the Δ 4-double bond. Dihydro-S1P did not affect stomatal apertures (Fig. 2b), providing evidence that the effect of S1P is specific and dependent on the presence of the Δ 4-double bond. Our results also indicate that dihydro-S1P can function as an inactive analogue in stomatal guard cells, an observation in good agreement with results obtained from animal cell studies¹¹.

In animal cells, the effect of S1P on cellular function is calcium dependent^{6–8}. As inclusion of 2 mM EGTA resulted in complete abolition of the effect of S1P on stomatal aperture, the same relationship holds in guard cells (Fig. 2c). To investigate the involvement of calcium directly, we pressure-microinjected the Ca²⁺-sensitive fluorescent indicator, Fura-2, into the cytosol of single guard cells and measured changes in [Ca²⁺]_{cyt} in response to 6 μ M (*n* = 4) and 50 μ M (*n* = 6) S1P. The resting levels of [Ca²⁺]_{cyt} in these cells (*n* = 10) were 50–120 nM. When challenged with 6 μ M S1P, oscillations in [Ca²⁺]_{cyt} were observed. The oscillations induced by 6 μ M S1P were rapid and asymmetrical, with a period of 3.8 ± 0.4 min (*n* = 4) (Fig. 3a). The elevations in [Ca²⁺]_{cyt}

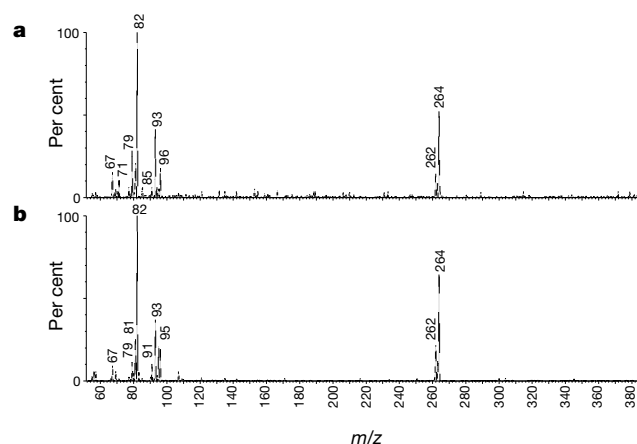


Figure 1 Positive-ion LC/MS/MS product ion spectra of *m/z* 380 (M+H)⁺. **a**, Product ion spectrum from a partially purified lipid extract of mature, fully expanded leaves of *C. communis* (see Methods). **b**, Product ion spectrum from authentic S1P.

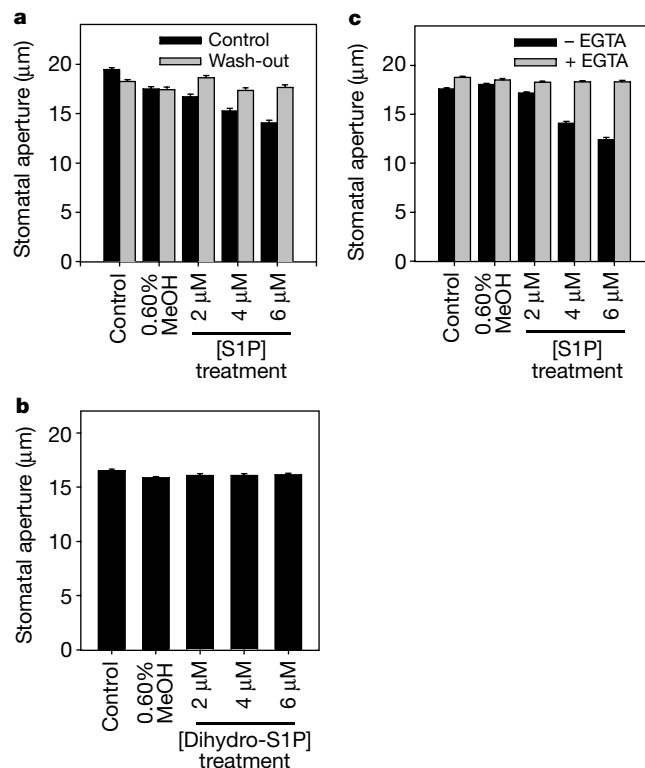


Figure 2 Effects of S1P and dihydro-S1P on stomatal aperture. **a**, Apertures of opened stomata after a 2-h incubation in S1P (black columns) and a subsequent 2-h wash-out (grey columns). **b**, Apertures of opened stomata after a 2-h incubation in dihydro-S1P. **c**, Apertures of opened stomata after a 2-h incubation in S1P (black bars) and S1P + 2 mM EGTA (grey columns). Values are means $\pm$ s.e.m. (*n* = 180).

ranged from 30 to 50 nM above resting levels. Rapid and symmetrical oscillations in $[Ca^{2+}]_{cyt}$ with a period of 2.8 ± 0.2 min were observed ($n = 6$) when guard cells were challenged with $50 \mu M$ S1P (Fig. 3b). The elevations in $[Ca^{2+}]_{cyt}$ induced by $50 \mu M$ S1P were 50–100 nM above resting levels. Although elevation in $[Ca^{2+}]_{cyt}$ is an integral downstream effect of S1P in all animal cell types examined^{6–8}, our data represent the first report of S1P-induced $[Ca^{2+}]_{cyt}$ oscillations. Challenging guard cells with S1P induced $[Ca^{2+}]_{cyt}$ oscillations in every experiment we performed.

To determine whether the effect of S1P on guard cell $[Ca^{2+}]_{cyt}$ is specific, we challenged single guard cells, pressure-microinjected with Fura-2 ($n = 5$) with $50 \mu M$ dihydro-S1P and measured $[Ca^{2+}]_{cyt}$. Resting levels of $[Ca^{2+}]_{cyt}$ in these cells were 80–120 nM, and when challenged with dihydro-S1P no change in $[Ca^{2+}]_{cyt}$ was observed ($n = 5$). When these same cells were subsequently challenged with 1 mM Ca^{2+} , a characteristic pattern of $[Ca^{2+}]_{cyt}$ oscillations¹² was observed, demonstrating that these cells possessed functional calcium homeostatic machinery (Fig. 3c). These results show that, like other signals^{13–16}, S1P induces closure of the stomatal pore by generating an increase in $[Ca^{2+}]_{cyt}$. In the case of S1P the increases take the form of oscillations in $[Ca^{2+}]_{cyt}$, which have a period and amplitude characteristic of the concentration of S1P used. There was a lag period of about 120 and 4 min after the initiation of the oscillations in response to $6 \mu M$ (Fig. 3d) and $50 \mu M$ (Fig. 3e) S1P, respectively, before changes in stomatal apertures were observed. These data point towards a physiological role for oscillations in $[Ca^{2+}]_{cyt}$ induced by S1P in mediating changes in guard cell turgor.

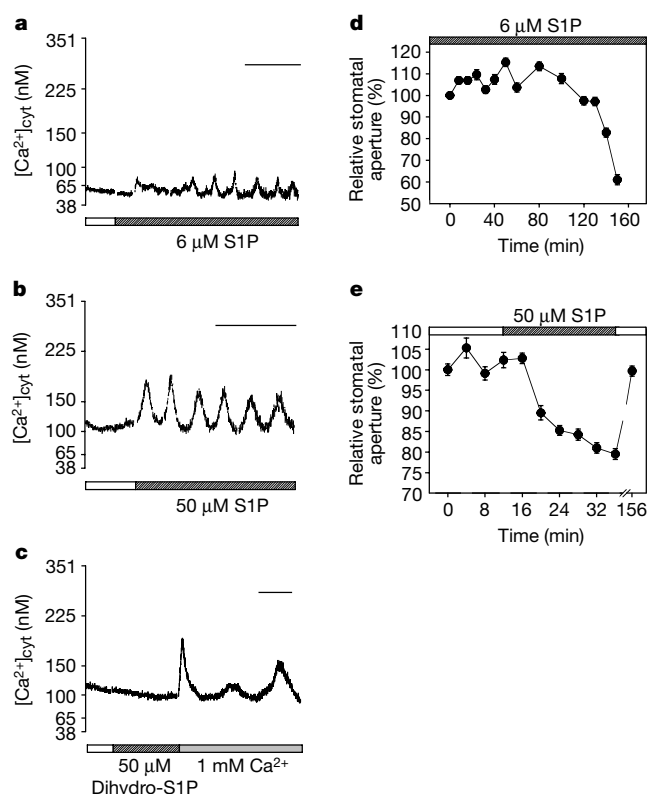


Figure 3 Effects of S1P and dihydro-S1P on guard cell $[Ca^{2+}]_{cyt}$ and kinetics of S1P-induced stomatal closure. **a**, Representative trace ($n = 4$) of oscillations in guard cell $[Ca^{2+}]_{cyt}$ induced by $6 \mu M$ S1P. **b**, Representative trace ($n = 6$) of oscillations in $[Ca^{2+}]_{cyt}$ induced by $50 \mu M$ S1P. **c**, Representative trace ($n = 5$) showing the inability of $50 \mu M$ dihydro-S1P to induce a change in $[Ca^{2+}]_{cyt}$. Time bar, 10 min. In **d** and **e** stomata were pre-opened in KCl-MES buffer for 3 h before transfer to KCl-MES buffer containing S1P. **d**, Rates of stomatal closure in response to $6 \mu M$ S1P. **e**, Rates of stomatal closure in response to $50 \mu M$ S1P. Values are means $\pm$ s.e.m. ($n = 40$).

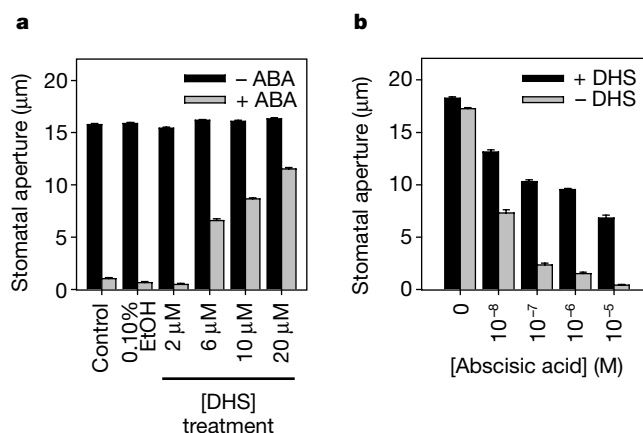


Figure 4 Effect of DHS on ABA-induced stomatal closure. **a**, Stomatal apertures after a 3-h pretreatment in buffer containing various concentrations of DHS (black columns), and 2 h after transfer to buffer containing 100 nM ABA (grey columns). **b**, Epidermal strips were pretreated with buffer only (grey columns) or buffer containing $20 \mu M$ DHS (black columns) for 3 h and apertures of opened stomata were measured after a subsequent 2-h incubation in buffer containing various concentrations of ABA. Values are means $\pm$ s.e.m. ($n = 180$).

As the regulation of stomatal turgor by abscisic acid (ABA) involves an increase in $[Ca^{2+}]_{cyt}$ (refs 17, 18), we next investigated whether S1P is a component of the guard cell ABA signalling pathway. We used DL-threo-dihydrosphingosine (DHS), which in animal cells is a competitive inhibitor of sphingosine kinase^{19,20}, to decrease cellular S1P levels. On its own, DHS did not affect stomatal apertures. However, in epidermal strips that had been pretreated with DHS, the effect of ABA on stomatal closure was attenuated. The attenuation of the ABA response was partial and dose dependent (Fig. 4a). In another experiment, we tested the ability of DHS to attenuate the effect of various concentrations of ABA ranging from 10^{-8} M to 10^{-5} M. Attenuation of the effect of ABA is partial, regardless of the concentration of ABA used (Fig. 4b). This is in agreement with the current understanding of an inherent redundancy in the ABA signal-transduction pathway regulating guard cell turgor⁵.

In summary, we have provided a positive identification of sphingosine-1-phosphate in plants. We also present evidence that the levels of S1P increase in leaves after drought and that S1P is a calcium-mobilizing compound active in the signal-transduction pathway linking ABA to reductions in stomatal aperture. Our data also suggest that the messenger role of S1P is evolutionarily conserved between animals and plants. □

Methods

Stomatal bioassay and $[Ca^{2+}]_{cyt}$ measurements

Commelina communis L. plants were grown from seed at 20 – $25^\circ C$, and stomatal aperture bioassays using isolated abaxial epidermal peels were conducted as described^{15–17}. We measured guard cell $[Ca^{2+}]_{cyt}$ as described¹².

Lipid extraction and analysis

Lipid was extracted from mature, fully expanded leaves. Freshly harvested leaves (20 g) were homogenized with liquid N_2 . The homogenate was extracted with 100 ml of ice-cold $1:2$ chloroform:methanol (v/v) for 45 min under constant agitation and filtered. We added 48 ml of chloroform, 48 ml of 1 M KCl and 2 ml of 10% NH_4OH to the filtrate and mixed it thoroughly. Phases were separated by centrifugation and the upper alkaline phase was retained, to which 80 ml of chloroform and 4 ml of concentrated HCl were added. The sample was mixed thoroughly, and the phases were separated by centrifugation. The resultant upper phase was discarded, and the lower chloroform phase was retained. The chloroform was evaporated to dryness, and the sample was kept at $-20^\circ C$ for analysis.

The extracted lipids were analysed by atmospheric pressure ionization (API) liquid chromatography (LC) tandem mass spectrometry (MS/MS). Both LC/MS and LC/MS/MS experiments were performed on a HP1050 liquid chromatograph with a quaternary solvent delivery system, coupled to a SCIEX (Thornhill, Ontario, Canada) API III+ triple quadrupole mass spectrometer equipped with an API source, through an IonSpray (ISP) interface. High-purity nitrogen was used as the curtain gas in the API source at a flow rate

of 0.81 min⁻¹. LC conditions were Vydac C₄ protein HPLC column (15 cm × 4.6 mm i.d.) using an acetonitrile:water:tetrahydrofuran gradient at 1 ml min⁻¹. For this interface, the column effluent was split using a zero-dead-volume T connector, with a split ratio of 25:1 determined by the length of a fused silica capillary on the waste side of the T. The remainder of the column effluent was fed to the transfer line of the interface. The ISP needle was maintained at 5.5 kV. High-purity nitrogen was used as the nebulizing gas at a flow rate of 0.61 min⁻¹. A Macintosh Quadra900 computer was used for instrument control, data acquisition and data processing. A dwell time of 10 ms Da⁻¹ was used for full-scan (*m/z* 50–400) LC/MS analyses.

MS/MS measurements were based on collision-induced dissociations (CID) within the radiofrequency-only quadrupole at a collision energy of 30 eV. Argon was used as the collision gas at an indicated thickness of 220. Extraction efficiency was obtained from replicate spiking experiments and was determined ~16%.

For drought experiments, water was withheld for a total of 11 days. Leaves were harvested at the end of the 11-day drought period. Parallel control samples were harvested from well-watered plants. A 11-day drought period resulted in a significant reduction in the fresh weight/dry weight ratio compared with well-watered controls (*P* < 0.001, *t* test). Lipids were extracted from leaf samples from well-watered plants and plants subjected to a drought treatment. SIP levels were quantified by LC/MS as described above. Estimation of SIP levels in the lipid extracts was based on a linear calibration (*r*² = 0.996) obtained from areas under peaks of LC/MS analysis of known concentrations of authentic SIP. SIP levels are expressed in terms of dry weight of leaves based on linear calibrations of fresh weight versus dry weight from well-watered control plants (*r*² = 1.000) and plants subjected to the drought treatment (*r*² = 0.999).

Preparation of sphingolipid solutions

SIP, dihydro-SIP and DHS were obtained from Biomol (Plymouth Meeting, Pennsylvania, USA) and stock solutions were prepared in methanol (SIP and dihydro-SIP) or ethanol (DHS).

Received 8 August 2000; accepted 11 January 2001

- Willmer, C. & Fricker, M. *Stomata* 2nd edn (Chapman & Hall, London, 1996).
- MacRobbie, E. A. C. Signal transduction and ion channels in guard cells. *Phil. Trans. R. Soc. Lond. B* **353**, 1475–1488 (1998).
- Sanders, D., Brownlee, C. & Harper, J. F. Communicating with calcium. *Plant Cell* **11**, 691–706 (1999).
- McAinsh, M. R. & Hetherington, A. M. Encoding specificity in Ca²⁺ signalling systems. *Trends Plant Sci.* **3**, 32–36 (1998).
- Blatt, M. R. Ca²⁺ signalling and control of guard-cell volume in stomatal movements. *Curr. Opin. Plant Biol.* **3**, 196–204 (2000).
- Mattie, M. E., Brooker, G. & Spiegel, S. Sphingosine-1-phosphate, a putative second messenger, mobilizes calcium from internal stores via an inositol trisphosphate-independent pathway. *J. Biol. Chem.* **269**, 3181–3188 (1994).
- Spiegel, S., Cuvillier, O., Fuio, E. & Milstein, S. in *Sphingolipid-mediated Signal Transduction* (ed. Hannun, Y. A.) 121–135 (Springer, Berlin, 1997).
- Pyne, S. & Pyne, N. J. Sphingosine 1-phosphate signalling in mammalian cells. *Biochem. J.* **349**, 385–402 (2000).
- Merrill, A. H. Jr & Sweeley, C. C. in *Biochemistry of Lipids, Lipoproteins and Membranes* (eds Vance, D. E. & Vance, J.) 309–339 (Elsevier Science, Amsterdam, 1996).
- Sullards, M. C., Lynch, D. V., Merrill, A. H. Jr & Adams, J. Structure determination of soybean and wheat glucosylceramides by tandem mass spectrometry. *J. Mass Spectrom.* **35**, 347–353 (2000).
- Berger, A., Bittman, R., Schmidt, R. R. & Spiegel, S. Structural requirements of sphingosylphosphocholine and sphingosine-1-phosphate for stimulation of Activator Protein-1 activity. *Mol. Pharmacol.* **50**, 451–457 (1996).
- McAinsh, M. R., Webb, A. A. R., Taylor, J. E. & Hetherington, A. M. Stimulus-induced oscillations in guard cell cytosolic free calcium. *Plant Cell* **7**, 1207–1219 (1995).
- Gilroy, S., Read, N. D. & Trewavas, A. J. Elevation of cytoplasmic calcium by caged calcium or caged inositol trisphosphate initiates stomatal closure. *Nature* **346**, 769–771 (1990).
- Blatt, M. R., Thiel, G. & Trentham, D. R. Reversible inactivation of K⁺ channels of *Vicia* stomatal guard cells following the photolysis of caged inositol 1,4,5-trisphosphate. *Nature* **346**, 766–769 (1990).
- Staxén, I. et al. Abscissic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. *Proc. Natl Acad. Sci. USA* **96**, 1779–1784 (1999).
- Leckie, C. P., McAinsh, M. R., Allen, G. J., Sanders, D. & Hetherington, A. M. Abscissic acid-induced stomatal closure mediated by cyclic ADP-ribose. *Proc. Natl Acad. Sci. USA* **95**, 15837–15842 (1998).
- McAinsh, M. R., Brownlee, C. & Hetherington, A. M. Abscissic acid-induced elevation of guard cell cytosolic Ca²⁺ precedes stomatal closure. *Nature* **343**, 186–188 (1990).
- Schroeder, J. I. & Hagiwara, S. Repetitive increases in cytosolic Ca²⁺ of guard cells by abscissic acid activation of nonselective Ca²⁺ permeable channels. *Proc. Natl Acad. Sci. USA* **87**, 9305–9309 (1990).
- Buehrer, B. M. & Bell, R. M. Inhibition of sphingosine kinase *in vitro* and in platelets. *J. Biol. Chem.* **267**, 3154–3159 (1992).
- Kohama, T. et al. Molecular cloning and functional characterization of murine sphingosine kinase. *J. Biol. Chem.* **273**, 23722–23728 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by a grant from the Westlakes Charitable Trust (Cumbria, UK). M.R.M. is grateful to The Royal Society for the award of a University Research Fellowship.

Correspondence and requests for materials should be addressed to A.M.H. (e-mail: a.hetherington@lancaster.ac.uk).

Hedgehog acts as a somatic stem cell factor in the *Drosophila* ovary

Yan Zhang & Daniel Kalderon

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

Secreted signalling molecules of the Hedgehog (Hh) family have many essential patterning roles during development of diverse organisms including *Drosophila* and humans^{1,2}. Although Hedgehog proteins most commonly affect cell fate, they can also stimulate cell proliferation. In humans several distinctive cancers, including basal-cell carcinoma, result from mutations that aberrantly activate Hh signal transduction³. In *Drosophila*, Hh directly stimulates proliferation of ovarian somatic cells^{4–6}. Here we show that Hh acts specifically on stem cells in the *Drosophila* ovary. These cells cannot proliferate as stem cells in the absence of Hh signalling, whereas excessive Hh signalling produces supernumerary stem cells. We deduce that Hh is a stem-cell factor and suggest that human cancers due to excessive Hh signalling might result from aberrant expansion of stem cell pools.

In adult *Drosophila* females, egg chambers are produced continuously in the germarium of each of the 15–18 ovarioles that are bundled together to form an ovary⁷. In regions 1 and 2a of the germarium, 16-cell germline cysts develop from germline stem cells (Fig. 1). Each cyst is enveloped by a monolayer of follicle cells in region 2b and separated from the next cyst by a short stalk as it buds from region 3 to form an egg chamber. Follicle and stalk cells derive from somatic stem cells that reside at the region 2a/2b border⁸. When a somatic stem cell divides, one daughter retains a stem cell identity and continues to divide as a stem cell for several days⁸. The other, 'pre-follicle cell' daughter proliferates for about eight cycles as its progeny associate with germline cysts, pass posteriorly down the ovariole over a 5–6-day period, and differentiate into multiple specialized cell types^{7,8}. Hedgehog (Hh) is expressed selectively in specialized non-proliferating, somatic 'terminal filament' and 'cap' cells at the anterior tip of the germarium⁴ (Fig. 1). Inactivation of Hh, using conditional *hh* alleles, arrests egg chamber budding, and causes germline cysts to accumulate in swollen germaria, suggesting that too few follicle cells are being produced⁴. Conversely, excessive Hh signalling in germarial region 2 can be induced by temporally controlled activation of an *hh* transgene or by inactivation of the Hh receptor Patched (Ptc), and causes marked overproliferation of somatic cells, which accumulate between egg chambers^{4–6} (Fig. 2).

We investigated the proliferative response to excessive Hh signal

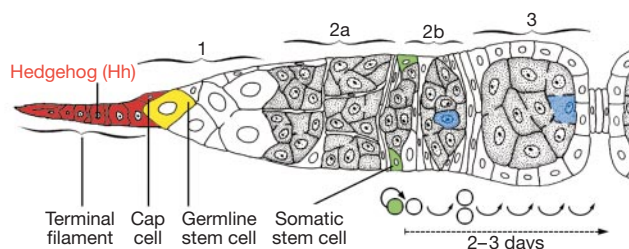


Figure 1 The *Drosophila* germarium. Regions 1 and 2a contain large proliferating germline stem cells (yellow), cystoblasts and cystocyte clusters (shaded), and smaller non-proliferating somatic cells. Somatic stem cells (green) renew and produce pre-follicle daughter cells (colourless), which divide roughly five times in the germarium (~50 h), three more times up to stage 6 (~30 h) and then arrest before differentiation, and exit from the ovariole (~30 h later). Hedgehog-producing cells and the oocyte are coloured red and blue respectively.

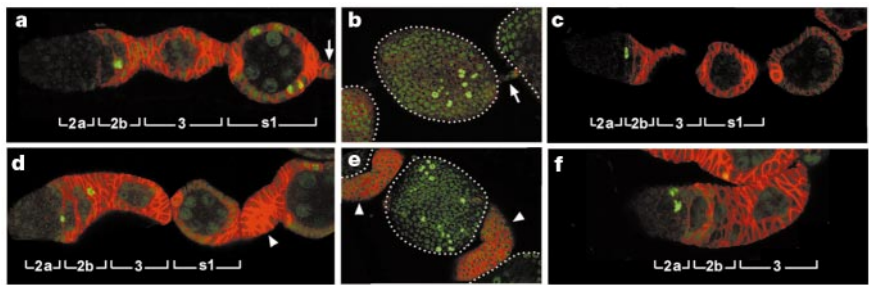


Figure 2 Phospho-histone H3 labelling of follicle cells and somatic stem cells. Wild-type ovarioles (**a–c**) and marked *ptc* mutant ovarioles (**d–f**) were labelled with antibody against phospho-histone H3 (green) and Fas III (red). The short stalks of wild-type ovarioles (arrows in **a** and **b**) are replaced by large masses of Fas III-positive cells in *ptc* mutant ovarioles (arrowheads in **d** and **e**). **a, d**, Mitotic cells (green) in region 2b also stain

with Fas III and are therefore not stem cells. **b, e**, Mitotic cells in the follicular monolayer around stage 6 egg chambers. **c, f**, Mitotic cells at the region 2a/2b border, anterior to the limits of Fas III staining, are putative stem cells. Ovarioles containing only *ptc* mutant cells were distinguished from mosaic ovarioles by the absence of *tub-lacZ* transgene products.

transduction further by using antibodies against phospho-histone H3, which stains cells only during mitosis⁹. We induced high levels of Hh signal transduction by inactivating *patched* (*ptc*) in marked somatic cell clones generated by heat-shock induced mitotic recombination¹⁰. Ovaries were examined 8 d later (Fig. 2) to ensure that all proliferating somatic cells assayed derived from stem cells that were present at the time of clone induction. As each ovariole contains more than one somatic stem cell⁸, this procedure generates some ovarioles containing only *ptc* mutant somatic cells and others that are mosaic for wild-type and *ptc* mutant cells. The number of mitotic somatic cells in germarial region 2 (Fig. 2a, d) in ovarioles containing only *ptc* mutant somatic cells was twice that in wild-type ovarioles (Table 1). A similar ratio was observed in region 3 of the germarium and in newly budded (stage 1–2) egg chambers (Table 1), suggesting an early increase in proliferation of *ptc* mutant cells followed by wild-type rates of proliferation of a twofold enlarged cell population. Accordingly, the follicular monolayers surrounding stage 6 egg chambers in wild-type and *ptc* mutant ovarioles (Fig. 2b, e) contained almost identical

numbers of mitotic cells within numerically equivalent cell populations (Table 1). In ovarioles mosaic for *ptc* mutant and wild-type somatic cells, the number of cells in mitosis was increased roughly 1.5-fold in region 2; again, this ratio was maintained in region 3 and in the earliest egg chambers. This is consistent with a cell-autonomous effect of *ptc* inactivation⁶, affecting roughly half of the somatic cells in a mosaic germarium.

No definitive positive marker for ovarian somatic stem cells has been described¹¹. However, when a single wild-type lineage of proliferating somatic cells is labelled by a stem cell recombination event, only one cell at the most anterior position fails to stain with Profilin¹¹ or Fasciclin III (see below), suggesting that this cell is a somatic stem cell. We therefore double-stained ovarioles with antibodies against Fasciclin III (Fas III) and phospho-histone H3, and counted the number of proliferating cells at the border of region 2a/2b that failed to stain with Fas III (Fig. 2c, f). The number of such putative stem cells in mitosis in pure *ptc* mutant ovarioles was roughly twice (0.17/0.08) that in wild-type ovarioles (Table 1). This shows that excessive Hh signal transduction promotes somatic stem cell proliferation and that this might account entirely for the ensuing somatic cell overproliferation.

To establish whether excessive Hh signalling accelerated somatic stem cell cycles or increased the number of stem cells in an ovariole, we counted stem cells by using mitotic recombination to provide up to four distinct genetic lineage labels (Fig. 3). We induced *ptc* mutant, or control, somatic stem cell clones in a genotype that allowed a loss of *tub-lacZ* and the gene for green fluorescent protein (*GFP*) transgenic markers by independent heat-shock-induced recombination events. We then waited 3 d for *ptc* inactivation to

Table 1 Number of mitotic somatic cells per ovariole

Ovarioles	Stem cell	Region 2b	Region 3	Stage 1–2	Stage 6*
Wild-type	0.08	0.6	0.8	1.4	11.7
All <i>ptc</i> [−]	0.17†	1.2‡	1.5‡	2.8‡	10.9
Mosaic <i>ptc</i> [−]	0.14	1.0	1.2	2.3	

Each number is the mean from 97 or more ovarioles.

* Only the monolayer of follicle cells surrounding stage 6 egg chambers was examined.

† *P* = 0.052 in comparison with wild-type.

‡ Significantly different from wild-type value (*P* < 0.01).

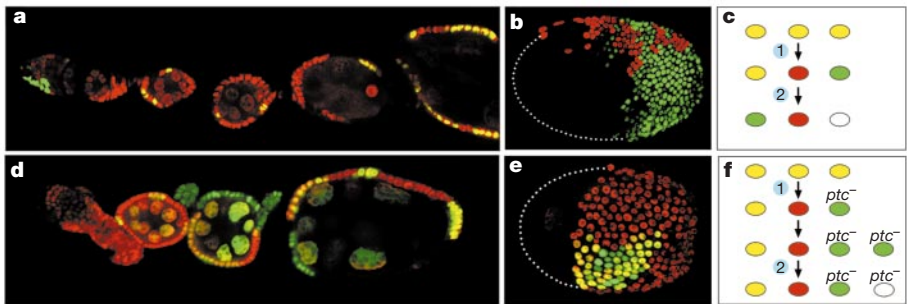


Figure 3 Counting somatic stem cells in wild-type and *ptc* mutant ovarioles. Stem cell clones generated in wild-type (**a–c**) and *ptc* heterozygous flies (**d–f**) stained for expression of *lacZ* (red) and *GFP* (green). Wild-type ovarioles (**a**) and some individual wild-type egg chambers (**b**) contained up to three distinguishable follicle cell types, whereas some *ptc* mutant ovarioles (**d**) or even single egg chambers (**e**) included four distinguishable follicle cell types. **c, f**, Examples of stem cell recombination events

induced by the indicated heat shocks (1 and 2) that generate three-colour wild-type ovarioles (**c**), and four-colour *ptc* mutant ovarioles (**f**) from three *lacZ*⁺ *GFP*⁺ (red plus green staining) stem cells. Germline cell nuclei are visible in some optical sections shown (green anterior cells in **a**, several in **d**) but are readily distinguished from follicle cells by their larger size.

Table 2 Distribution of distinguishable somatic stem cell genotypes in wild-type and *ptc* mutant ovarioles

Ovarioles	Distinguishable genotypes	8 d†	10 d	12 d	14 d	16 d*
Wild-type	1	16	21	37	41	48
	2	57	53	57	52	51
	3	27	26	6	7	1
<i>ptc</i>	1	14	21	24	35	65
	2	30	47	64	61	29
	3	42	29	9	4	6
	4	14	3	3	0	0

Each entry is expressed as the percentage of ovarioles with the stated number of distinguishable genotypes. At least 100 ovarioles were counted for each entry.

* Distribution of ovarioles significantly different between wild-type and *ptc* by χ^2 test ($P < 0.05$).

† Random recombination events (with overall efficiency of 73% over two heat shocks) predict a distribution of 20% (one genotype), 60% (two genotypes), 25% (three genotypes) for three stem cells.

affect stem cell behaviour and heat-shocked animals again to ensure that all stem cells, including any newly engendered by loss of *ptc*, could undergo recombination events distinguishing them from other stem cells in the same ovariole. After 8 d, the number of distinguishable somatic cell genotypes (*lacZ*⁺ *GFP*⁺, *lacZ*⁺ *GFP*⁺, *lacZ*⁺ *GFP*⁺ and *lacZ*⁺ *GFP*⁺) in a single ovariole was scored (Fig. 3). This provides a minimum measure of the number of somatic stem cells present at the time of the second heat shock. Wild-type ovarioles contained up to three distinguishable genotypes, implying the presence of up to three somatic stem cells in a single ovariole (Fig. 3a–c). Indeed, the observed distribution of marked genotypes indicates that most or all wild-type ovarioles contain three somatic stem cells (Table 2). Four distinguishable somatic cell genotypes were never observed in wild-type ovarioles but were present in 14% of *ptc* mutant ovarioles (Fig. 3d–f; Table 2), showing that the latter

contained at least four somatic stem cells. The constraints of the experiment require that these ‘four stem cell ovarioles’ contain two wild-type stem cells and two *ptc* mutant stem cells (Fig. 3f). We therefore deduce that such ovarioles originally included a single *ptc* mutant stem cell that duplicated within 3 d to form two stem cells. We are confident that the different follicle cell genotypes observed do indeed derive from recombination events in stem cells (and therefore reflect stem cell number) because four genotypes were evident in some *ptc* mutant ovarioles as long as 12 d after heat-shock (Table 2) and spontaneous recombination events in these flies were infrequent.

It has previously been estimated that there are two, rather than three, somatic stem cells in a wild-type ovariole, by counting the progeny of single marked stem cells as constituting roughly 50% of the follicle cells in an ovariole⁸. There are several possible explanations for this discrepancy. Later, we present further evidence that in our studies wild-type ovarioles contained three stem cells.

To reveal the effects of *ptc* inactivation on stem cell number directly we positively marked single wild-type and *ptc* mutant stem cell lineages with *lacZ* activity^{8,12} (Fig. 4b), and we counter-stained with antibody against Fas III (Fig. 4). We used a very mild heat-shock treatment (37 °C for 10 min) to induce mitotic recombination so that marked ovarioles generally contained only a single *lacZ*⁺ clone, and we examined ovaries 5 d later to ensure that clones were of stem cell origin. In wild-type ovarioles, most *lacZ*⁺ clones (32/46) included only a single cell at the region 2a/2b border that did not stain with Fas III (Fig. 4c); only a single clone included two such cells (Table 3). This is consistent with the idea that *lacZ*⁺ *Fas III*⁺ somatic cells at the region 2a/2b border are indeed stem cells. In *ptc* heterozygous animals, some recombinant *lacZ*⁺ stem cell clones also lost *ptc* activity through a second recombination event (Fig. 4b). These *ptc* mutant stem cell clones were identified by their over-proliferation and accumulation between egg chambers. Most of the *ptc* mutant clones (31/41) included two *lacZ*⁺ *Fas III*⁺ cells at the region 2a/2b border (Fig. 4d, f); only one *ptc* mutant clone included a single such putative stem cell (Table 3). The two *ptc* mutant stem cells must, in most cases, have derived from a single recombination event that initially generated a single *ptc* mutant stem cell. Thus, in almost every ovariole examined, a loss of *ptc* activity led to a

Table 3 Numbers of wild-type and *ptc* mutant lineages with zero, one and two stem cells

Ovarioles	Two stem cells	One stem cell	No stem cells
Wild-type	1	32	13
<i>ptc</i>	31	1	9

A small number of ovarioles could not be scored definitively because some somatic cells in region 1 and early region 2 (generally described as non-proliferating) expressed β -galactosidase.

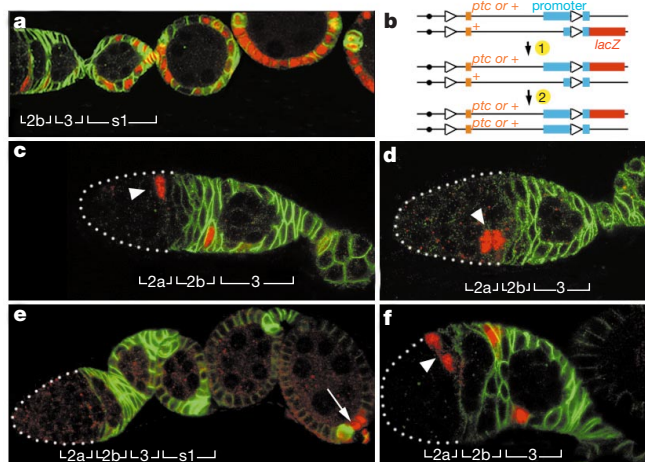


Figure 4 Direct labelling of wild-type and *ptc* mutant somatic stem cells. Ovarioles were stained for *lacZ* expression (red) and Fas III (green) 5 d after induction of mitotic recombination in *ptc* heterozygotes or control flies by a single heat shock. **a**, In control ovarioles *lacZ*⁺ clones are induced by event 1 depicted in **b** and populated both a fraction of the germarium and the epithelium of several egg chambers from stage 1–6. **c**, Wild-type stem cell clones generally included a single *lacZ*⁺ *Fas III*⁺ cell at the region 2a/2b border. **e**, Occasionally no such putative stem cell was present. **b**, In *ptc* heterozygotes the generation of marked *ptc* mutant clones requires two mitotic recombination events (1 and 2) in the same cell. **d, f**, Such clones, identified by *lacZ*⁺ cells in enlarged stalks (not shown), almost invariably included two *lacZ*⁺ *Fas III*⁺ putative stem cells, either adjacent (**d**) or not (**f**; greater separations were also seen).

Table 4 Percentage of *lacZ*⁺, mosaic and *lacZ*[−] ovarioles in wild-type, *smo* and *ptc* heterozygotes

Ovarioles	Genotype	Days after heat shock				
		3	8	10	12	14
Wild-type	<i>lacZ</i> ⁺	0	44	51	61	60
	Mosaic	100	56	49	36	32
	<i>lacZ</i> [−]	0	0	2	3	8
<i>smo</i>	<i>lacZ</i> ⁺	0	87	94	100	100
	Mosaic	100	13	6	0	0
	<i>lacZ</i> [−] (<i>smo</i> [−])	0	0	0	0	0
<i>ptc</i>	<i>lacZ</i> ⁺	0	32	33	33	34
	Mosaic	100	57	51	39	31
	<i>lacZ</i> [−] (<i>ptc</i> [−])	0	11	15	29	35

Each entry derives from at least 100 ovarioles of adults after heat-shock-induced mitotic recombination in *lacZ*⁺/*lacZ*[−] third-instar larvae.

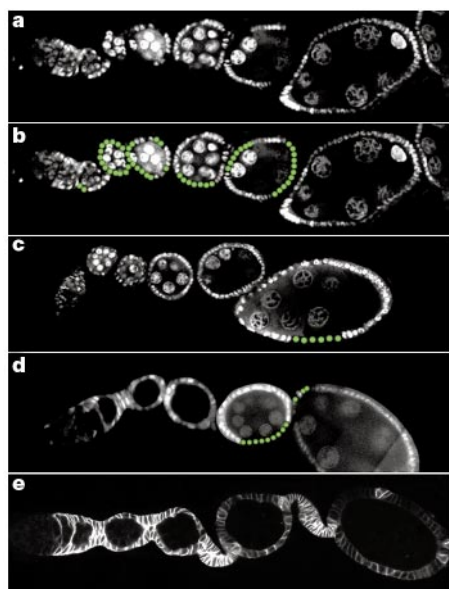


Figure 5 Somatic stem cell proliferation requires *smo* and *ci* activities. **a–d**, Wild-type clones (**a**, **b**) *smo* mutant clones (**c**) and *ci* mutant clones (**d**) were marked by loss of a *tub-lacZ* transgene and were identified 8 d later by the absence of *lacZ* expression (colourless nuclei, artificially coloured green for clarity in **b–d**). Panels **a** and **b** show the same ovariole. **e**, Fas III staining highlights excessive cells between egg chambers resulting from the expression of an activated *ci* transgene initiated 8 d earlier.

cell-autonomous doubling of somatic stem cell number. In 19 of these instances, the two *ptc* mutant stem cells were directly adjacent (Fig. 4d), suggesting that excessive Hh signal transduction allowed local expansion of a stem cell niche. In the remaining 12 cases the additional *ptc* mutant stem cell had migrated away from its presumed sister cell (Fig. 4f). In several wild-type and *ptc* mutant ovarioles, no stem cell was visible for a *lacZ*⁺ lineage (Fig. 4e; Table 3). We presume that in these cases the marked stem cell died or started to proliferate like a pre-follicle cell within the 5-d period between clone induction and ovary dissection.

We investigated whether Hh is required for somatic stem cell maintenance or proliferation by using conditional *hh* alleles and by generating somatic cell clones lacking *smoothed* (*smo*) activity. Inactivation of *smo* universally blocks Hh signal transduction cell-autonomously¹. Control and *smo* mutant clones, each marked by loss of a *tub-lacZ* transgene, were examined at various times after induction. At 3 d, *smo* mutant clones of the same size as control clones were efficiently recovered in egg chambers of all stages (Table 4). However, by 8 d, when only stem cell clones should remain in the ovariole, very few *smo* mutant clones were observed. The *smo* mutant clones that were present generally occupied only a portion of a single egg chamber (Fig. 5c), in contrast to a typical control stem cell clone, which spread throughout the ovariole (Fig. 5a, b). By 12 d after clone induction no *smo* mutant clones were observed (Table 4). These results demonstrate that a cell that is unable to transduce a Hh signal cannot proliferate as a somatic stem

Table 6 Percentage of mosaic ovarioles induced in wild-type adults of different ages

Age (d)	0	2	4	6	8	10	12
Mosaic (%)	83	83	84	85	87	85	81

Clones were induced in *lacZ*⁺/*lacZ*⁻ adults at the given times after eclosion and the percentage of mosaic ovarioles was scored 8 d later in a sample of at least 100 ovarioles.

cell. We cannot be certain of the fate of such cells, but we suspect that *smo* mutant somatic stem cells remain abnormally quiescent for up to 7–8 d and at some point during this period acquire the characteristics of a pre-follicle cell, proliferating normally in that capacity to produce a clone occupying roughly one-third of an egg chamber.

In *hh*^{ts2} animals shifted to the restrictive temperature of 29 °C a statistically significant reduction in the number of mitotic cells labelled by phospho-histone H3 antibody was seen in region 2b by 72 h, and in germarial region 3 by 96 h (Table 5). Double-staining with Fas III antibody showed that the number of stem cells undergoing mitosis was already reduced by 48 h and was undetectable by 72 h. These reductions were not statistically significant because wild-type ovarioles sampled at these time points collectively contained only two or three somatic stem cells. Nevertheless, these mitotic labelling data are the first direct evidence that Hh inactivation reduces somatic cell proliferation and are consistent with a primary effect on stem cell proliferation.

The effects of Hh signalling on cell fate determination in *Drosophila* are mediated largely by altering the activity of the transcription factor Cubitus interruptus (Ci)¹. We examined the role of Ci in somatic stem cell proliferation first by inducing somatic clones lacking *ci* activity. As with *smo*, very few clones were recovered 8–10 d after clone induction and these clones occupied only a small proportion of the ovariole (Fig. 5d), indicating that stem cells cannot proliferate normally in the absence of *ci* activity. When we induced the expression of a constitutively active derivative of Ci (ref. 13) by heat-shock-induced excision of a transcriptional terminator, we recovered ovarioles showing massive overproliferation of somatic cells, which accumulated between egg chambers (Fig. 5e) as observed for *ptc* mutant ovarioles. Thus, the activity of Hh as a stem cell factor seems to depend on Ci-mediated regulation of transcription. The definition of relevant target genes presents an avenue into understanding how stem cell identity and proliferation are regulated.

Somatic stem cells seem to have a limited lifespan⁸. We investigated the longevity of wild-type ovarian somatic stem cells by inducing marked clones in third-instar larvae and examining adult ovaries 8–14 d later. The proportion of mosaic ovarioles declined from 56% (in 3-day-old adults) to 32% (in 9-day-old adults) during this period (Table 4), indicating a loss of some of the stem cells present at the time of clone induction. To see whether these stem cells were being replaced we performed a complementary experiment in which clones were induced in adults of various ages and then assayed 8 d later. The frequency of mosaic ovarioles was unchanged in adults from zero to 12 d after eclosion (Table 6), suggesting that the total number of somatic stem cells was constant over this period. We therefore deduce that, on average, wild-type stem cells are lost roughly every 8–9 d but are subsequently

Table 5 Number of mitotic somatic cells per ovariole*

Days at 29 °C	+/+					<i>hh</i> ^{ts} / <i>hh</i> ^{ts}					<i>hh</i> ^{ts} /+				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Stem cell	0.02	0.06	0.04	0.06	0.06	0.02	0.08	0.02	0.00	0.00	0.02	0.06	0.04	0.02	0.02
Region 2b	0.46	0.60	0.60	0.54	0.44	0.36	0.66	0.36	0.16*	0.22	0.62	0.74	0.46	0.36	0.30
Region 3	0.66	1.28	1.36	1.08	0.98	0.70	0.94	0.94	0.77	0.48*	0.82	1.12	0.98	0.72	0.58
Stage 1–2	2.51	3.12	4.22	3.76	3.90	2.68	2.42*	2.72*	2.14*	1.30*	2.44	2.58	2.98	2.76	1.92

Each number is the mean from 50 ovarioles, scored every 24 h after transfer of 2-day-old adults from permissive (18 °C) to restrictive (29 °C) temperature.

* Significantly different from +/+ values (*P* < 0.05).

replaced by division of another stem cell in the same ovariole to produce two sister stem cells.

We also note that the distribution of genotypes in wild-type ovarioles at day 8 (Table 4), before significant stem cell loss and replacement, is consistent with the presence of three somatic stem cells. A simple model of random, independent conversion of each of three stem cells to a *lacZ*⁺ genotype with probability $P = 0.5$ predicts a distribution (*lacZ*⁺ 42%; mosaic 56%; *lacZ*[−] 2%) similar to that observed. This model would predict a much higher proportion of *lacZ*[−] ovarioles relative to mosaic ovarioles for all P values if only two somatic stem cells were present.

When marked *ptc* mutant *lacZ*[−] stem cell clones were induced in third-instar larvae, the number of ovarioles containing only *lacZ*⁺ cells remained roughly constant from 8 to 14 d later (Table 4), indicating that *ptc* mutant stem cells were not readily lost from mosaic ovarioles. Conversely, there was a strong preference for mosaic ovarioles to become populated entirely by *ptc* mutant cells over time. For example, 14 d after clone induction, 35% of ovarioles were composed entirely of *ptc* mutant *lacZ*[−] clones, whereas in wild-type animals only 8% of ovarioles contained solely *lacZ*[−] somatic cells (Table 4). The stability and apparently aggressive behaviour of *ptc* mutant clones is likely to be due in some part to the ability of *ptc* mutant stem cells to duplicate soon after they arrive in a vacated stem cell niche. However, it is possible that *ptc* mutant stem cells also actively displace wild-type stem cells from their niches.

All of our observations can most easily be rationalized by postulating that there are only a limited number of positions (literally three) within a germarium where somatic cells receive the correct combination of extracellular signals to instruct them to behave as stem cells. In a wild-type ovariole one of the two daughters of a stem cell division necessarily leaves this environment and therefore does not adopt a stem cell fate. Occasionally, a stem cell might leave its niche and proliferate as a pre-follicle cell. The vacant niche becomes occupied shortly afterwards by the daughter of another stem cell. In this model the division of stem cells is not intrinsically asymmetric. Rather, the different behaviour of the two daughters is dictated by their physical environment, so that both daughters can become stem cells if two stem cell niches are vacant. Hh seems to be one of the key environmental signals for directing somatic stem cell behaviour. Proliferation of a stem cell requires Hh signal transduction; excessive Hh signal transduction allows expansion of the stem cell niche. Our results suggest that each stem cell niche can accommodate two stem cells if these cells are undergoing excessive Hh signal transduction.

Because Hh signalling, like all major signalling pathways, is used in a large number of contexts, there is no compelling reason to expect that the proliferative effects of Hh proteins are always targeted to stem cells. However, several characteristics of normal and deregulated development of vertebrate epidermis and hair follicles prompt us to suggest that Sonic hedgehog (Shh) might also act on stem cells in this context. Shh is required for hair follicle development in mice, and *Shh* mutant phenotypes include reduced proliferation of ectodermal stem cell derivatives^{14,15}. Conversely, inappropriate mutational activation of the Shh signal transduction pathway is universally observed in human familial and sporadic basal-cell carcinomas (BCCs)^{16–20}. Several observations suggest that activation of this pathway might be obligatory in the aetiology of BCC and can suffice to initiate this cancer^{16,21–24}. The cellular origin of BCC has not been defined, but outer root sheath cells of hair follicles, which include epidermal and follicular stem cells in a specialized bulged region^{25–27}, are strongly implicated²⁸. In the light of our present studies, we suggest that proliferation of follicular stem cells might normally require Shh, among other factors, and that mutational activation of the Shh signalling pathway in BCC acts specifically on these stem cells to increase their cycling frequency or their cell number, or perhaps even to expand their niche into or towards the surrounding epithelium. Any of these changes could

produce a sustainable large increase in a population of transiently proliferating cells, which would constitute the bulk of the carcinoma. At present these hypotheses concerning Hh-associated cancers are speculative and remain to be tested. □

Methods

Mutant *ptc* and *smo* clones marked by the absence of a *tub-lacZ* transgene were generated as described previously⁵ by using a standard heat shock of 60 min at 39 °C. To count stem cells, clones were induced in flies of genotype *hsp70-flp/+; FRT42D ptc*⁵² [or +]/*FRT42D P[tub-lacZ]; FRT82B P[hsp70-myc-GFP]/FRT82B*. To mark stem cell lineages with *lacZ*⁺ we used flies of genotype *hsp70-flp/+; FRT42D ptc*⁵² [or +] *X15-29/FRT42D X15-31* (ref. 12). Clones lacking *ci* activity and marked by the loss of GFP were generated by using flies of genotype *hsp70-flp/hsp70-flp; FRT42D P[Ci⁺] P[hsp70-GFP]/FRT42D; ci⁹⁴/ci⁹⁴* (ref. 29). Clones expressing activated *Ci* were generated by excision of an FRT-flanked *CD2* gene from an *act5C > CD2 > GAL4* transgene²⁹ in the presence of the activated *UAS-Ci5m* transgene from line T5m-s1 (ref. 13). Ovaries were stained as described previously for β -galactosidase, Fas III and phospho-histone H3 epitope⁶, except that secondary antibodies for double-staining used Alexa fluorescence probes (Molecular Probes). GFP encoded by an *hsp70-myc-GFP* transgene³⁰ was detected by direct fluorescent imaging after a heat shock (60 min, 39 °C) and a 3-h recovery period.

Received 27 October 2000; accepted 8 January 2001.

- Ingham, P. W. Transducing Hedgehog: the story so far. *EMBO J.* **17**, 3505–3511 (1998).
- Ruiz i Altaba, A. Gli proteins and Hedgehog signaling. *Trends Genet.* **15**, 418–425 (1999).
- Wicking, C., Smyth, I. & Bale, A. The hedgehog signalling pathway in tumorigenesis and development. *Oncogene* **18**, 7844–7851 (1999).
- Forbes, A. J., Lin, H., Ingham, P. W. & Spradling, A. C. Hedgehog is required for the proliferation and specification of ovarian somatic cells prior to egg chamber formation in *Drosophila*. *Development* **122**, 1125–1135 (1996).
- Forbes, A. J., Spradling, A. C., Ingham, P. W. & Lin, H. The role of segment polarity genes during early oogenesis in *Drosophila*. *Development* **122**, 3283–3294 (1996).
- Zhang, Y. & Kalderon, D. Regulation of cell proliferation and patterning in *Drosophila* oogenesis by Hedgehog signaling. *Development* **127**, 2165–2176 (2000).
- Spradling, A. C. Developmental genetics of oogenesis. In *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 1–70 (Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1993).
- Margolis, J. & Spradling, A. C. Identification and behavior of epithelial stem cells in the *Drosophila* ovary. *Development* **121**, 3797–3807 (1995).
- Hendzel, M. J. et al. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosome* **106**, 348–360 (1997).
- Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
- Spradling, A. C. et al. The *Drosophila* germarium: stem cells, germ line cysts and oocytes. *Cold Spring Harbor Symp. Quant. Biol.* **62**, 25–34 (1997).
- Harrison, D. & Perrimon, N. A simple and efficient method for generation of marked clones in *Drosophila*. *Curr. Biol.* **3**, 424–433 (1993).
- Price, M. A. & Kalderon, D. Proteolysis of Cubitus interruptus in *Drosophila* requires phosphorylation by protein kinase A. *Development* **126**, 4331–4339 (1999).
- Chiang, C. et al. Essential role for Sonic hedgehog during hair follicle morphogenesis. *Dev. Biol.* **205**, 1–9 (1999).
- St-Jacques, B. et al. Sonic hedgehog signaling is essential for hair development. *Curr. Biol.* **8**, 1058–1068 (1998).
- Xie, J. et al. Activating smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90–92 (1998).
- Dahmane, N., Lee, J., Robins, P., Heller, P. & Ruiz i Altaba, A. Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours. *Nature* **389**, 876–881 (1997).
- Gailani, M. R. et al. The role of the human homolog of *Drosophila* patched in sporadic basal-cell carcinomas. *Nature Genet.* **14**, 78–81 (1996).
- Hahn, H. et al. Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* **85**, 841–851 (1996).
- Johnson, R. L. et al. Human homolog of patched, a candidate gene for the basal-cell nevus syndrome. *Science* **272**, 1668–1671 (1996).
- Fan, H. R., Oro, A. E., Scott, M. P. & Khavari, P. A. Induction of basal-cell carcinoma features in transgenic human skin expressing sonic hedgehog. *Nature Med.* **3**, 788–792 (1997).
- Grachtchouk, M. et al. Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nature Genet.* **24**, 216–217 (2000).
- Nilsson, M. et al. Induction of basal cell carcinomas and trichoepitheliomas in mice overexpressing Gli-1. *Proc. Natl Acad. Sci. USA* **97**, 3438–3443 (2000).
- Oro, A. E. et al. Basal cell carcinomas in mice overexpressing Sonic hedgehog. *Science* **276**, 817–821 (1997).
- Hardy, M. H. The secret life of the hair follicle. *Trends Genet.* **8**, 55–61 (1992).
- Rochat, A., Kobayashi, K. & Barrandon, Y. Location of stem cells of human hair follicles by clonal analysis. *Cell* **76**, 1063–1073 (1994).
- Taylor, G., Lehnert, M. S., Jensen, P. J., Sun, T.-T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
- Kruger, K., Blume-Peytavi, U. & Orfanos, C. E. Basal cell carcinoma possibly originates from the outer root sheath and/or the bulge region of the vellus hair follicle. *Arch. Dermatol. Res.* **291**, 253–259 (1999).
- Jiang, J. & Struhl, G. Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493–496 (1998).
- Methot, N. & Basler, K. Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819–831 (1999).

Acknowledgements

We thank D. Harrison, J. Jiang, N. Methot, N. Patel and A. Spradling for reagents; J. Briscoe for help with confocal microscopy; D. Rabinowitz for statistical modelling of stem cell behaviour; J. Erickson, J. Mohler, M. A. Price and R. Lavker for critical discussions. This work was supported by an NIH grant to D.K.

Correspondence and requests for materials should be addressed to D.K. (e-mail: ddk1@columbia.edu).

Structural basis for co-stimulation by the human CTLA-4/B7-2 complex

Jean-Claude D. Schwartz^{*†}, Xuewu Zhang^{†‡}, Alexander A. Fedorov[§], Stanley G. Nathenson^{*‡} & Steven C. Almo[§]

^{*} Department of Microbiology and Immunology, [‡] Department of Cell Biology, and [§] Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

[†] These authors contributed equally to this work

Regulation of T-cell activity is dependent on antigen-independent co-stimulatory signals provided by the disulphide-linked homodimeric T-cell surface receptors, CD28 and CTLA-4 (ref. 1). Engagement of CD28 with B7-1 and B7-2 ligands on antigen-presenting cells (APCs) provides a stimulatory signal for T-cell activation, whereas subsequent engagement of CTLA-4 with these same ligands results in attenuation of the response¹. Given their central function in immune modulation, CTLA-4- and CD28-associated signalling pathways are primary therapeutic targets for preventing autoimmune disease, graft versus host disease, graft rejection and promoting tumour immunity^{1,2}. However, little is known about the cell-surface organization of these receptor/ligand complexes and the structural basis for signal transduction. Here we report the 3.2-Å resolution structure of the complex between the disulphide-linked homodimer of human CTLA-4 and the receptor-binding domain of human B7-2. The unusual dimerization properties of both CTLA-4 and B7-2 place their respective ligand-binding sites distal to the dimer interface in each molecule and promote the formation of an alternating arrangement of bivalent CTLA-4 and B7-2 dimers that extends throughout the crystal. Direct observation of this CTLA-4/B7-2 network provides a model for the periodic organization of these molecules within the immunological synapse and suggests a distinct mechanism for signalling by dimeric cell-surface receptors.

The specificity of the T-cell response depends on the engagement of the T-cell receptor (TCR) with the cognate major histocompatibility complex (MHC)-peptide complex on the surface of APCs³. Effective regulation of this response requires a number of antigen-independent co-stimulatory signals, the most important of which are provided by the related T-cell surface molecules, CTLA-4 and CD28 (approximately 30% identity)¹. Studies indicate that CTLA-4 both disrupts stimulatory signalling complexes by competing with CD28 for binding the B7 isoforms, and promotes the assembly of inhibitory signalling complexes⁴. To examine the structural basis for co-stimulation, we have solved the structure of the complex formed by the disulphide-linked homodimer of human CTLA-4 and the receptor-binding domain of human B7-2. The asymmetric unit contains two monomers of each CTLA-4 and B7-2, providing two independent views of the CTLA-4/B7-2 interface. Notably, crystal packing generates an alternating network of bivalent CTLA-4 and B7-2 homodimers. The CTLA-4 and B7-2 monomers are both two-layer β -sandwiches that display the chain topology characteristic of the immunoglobulin variable (V-type) domains present in antigen

receptors such as the TCR β -chains and antibody V_L and V_H domains⁵. The front and back sheets of CTLA-4 contain strands A'GFCC' and ABEDC', respectively, whereas the front and back sheets of B7-2 are composed of AGFCC'C'' and BED strands, respectively. The individual CTLA-4 monomers in the present crystal structure are highly similar to the monomers observed in the murine CTLA-4 crystal structure⁶ and the human CTLA-4 NMR structure⁷, with root mean square (r.m.s.) deviations of 1.6 Å and 2.2 Å, respectively. Furthermore, the structure of the B7-2 monomer is similar to that observed for both unliganded B7-1 (ref. 8) and B7-2 (X.Z. and J.-C.D.S., unpublished results), displaying r.m.s. deviations of 2.3 Å and 0.8 Å, respectively.

The CTLA-4/B7-2 binding interface is formed by packing interactions between the front sheet of each molecule (Fig. 1a). CTLA-4 makes extensive contribution of residues from the FG loop (the complementary determining region 3 (CDR3)-like segment) and to a lesser extent the C and C' strands. These segments pack against the G, F, C, C' and C'' strands of B7-2, which form a shallow concave surface as a consequence of the natural twist of the sheet. The two independent CTLA-4/B7-2 interfaces have very similar atomic contacts, and bury 1,290-Å² (interface A) and 1,212 Å² (interface B) of total accessible surface area. The size of this interface compares favourably with other complexes of V-type immunoglobulin domains, such as the heterophilic CD2/CD58 complex⁹ and homophilic E-cadherin adhesion dimer¹⁰ that bury 1,160 Å² and 1,300 Å², respectively. Although the CD2/CD58 and E-cadherin interfaces are formed by packing of the strands residing on the front sheet of each interacting molecule, the CTLA-4/B7-2 interface is formed by the β -strands of B7-2 and surface loops on CTLA-4. In the crystal, the two CTLA-4/B7-2 interfaces show a difference in geometric complementarity, with shape correlation statistics (Sc values) of 0.66 and 0.58 for the A and B interfaces, respectively (see Methods). The differences in buried surface area and surface complementarity observed in the two CTLA-4/B7-2 interfaces arise from an approximate 13° rotational difference between the two CTLA-4 molecules relative to their respective B7-2 partners (see Methods).

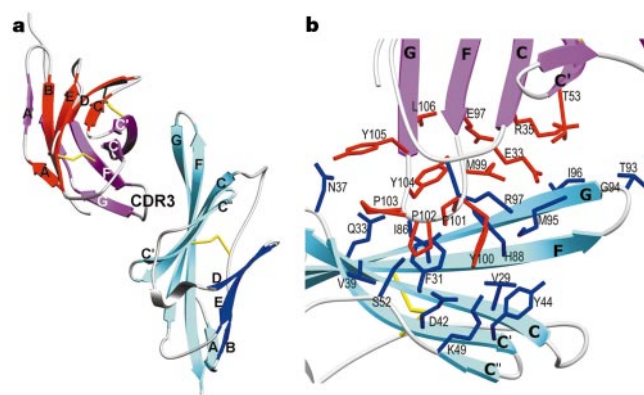


Figure 1 The CTLA-4/B7-2-binding interface. **a**, Ribbon diagram of the CTLA-4/B7-2 monomers that form the binding interface. For CTLA-4, the strands are labelled A (residues 4–8), A' (residues 11–13), B (residues 19–26), C (residues 33–42), C' (residues 45–54), C'' (residues 56–60), D (residues 67–72), E (residues 75–81), F (residues 90–99) and G (residues 105–109; 112–115). The A'GFCC' (front) and ABEDC' (back) sheets are coloured pink and red, respectively. The CDR3-like loop of CTLA-4 is labelled. For B7-2, the strands are labelled A (residues 3–7), B (residues 9–11), C (residues 27–35), C' (residues 38–45), C'' (residues 47–49), D (residues 60–64), E (residues 67–73), F (residues 81–91) and G (residues 94–108). The AGFCC'C'' (front) and BED (back) sheets are coloured dark blue and blue, respectively. All inter-sheet disulphides are shown in yellow. **b**, Detailed view of the human CTLA-4/B7-2 interface. The interface is formed by residues from the front sheets of CTLA-4 (CDR3 and the C and C' strands) and a concave surface on B7-2 (the G, F, C, C' and C'' strands and the CC', C'D and FG loops).

All subsequent discussions will refer to interface A.

The CTLA-4/B7-2 interface is stabilized by five hydrogen bonds and 28 van der Waals contacts involving a number of conserved residues, including the Met-Tyr-Pro-Pro-Pro-Tyr-Tyr sequence (residues 99–105) from CTLA-4 that corresponds to CDR3. As also observed in the murine CTLA-4 crystal structure⁶, the three consecutive proline residues in the complex adopt an unusual *cis-trans-cis* conformation, which promotes numerous van der Waals interactions with Phe 31, Gln 33, Asn 37, Leu 38, Val 39, Glu 42 and Ser 52 from the C and C' strands, and the CC' and C'D loops of B7-2 (Fig. 1a, b). The three tyrosine residues flanking these prolines provide additional van der Waals contacts with Val 29, Phe 31, Gln 33, Tyr 44, Ile 86, His 88 and Met 95, and participate in a series of potential hydrogen bonds, involving the side-chain phenolic hydroxyls, with Asn 37, Glu 42, Lys 49 and Arg 97 of B7-2. Additional contacts involve Glu 33 and Arg 35 (C strand), Thr 53 (C' strand) and Met 99 (CDR3) of CTLA-4 with residues 93–97 on the G strand of B7-2. An ionic interaction between Glu 97 of CTLA-4 and Arg 97 of B7-2 also contributes to the interface. These contacts are consistent with mutagenesis data, indicating that residues in CDR3 and the C strand of CTLA-4 (refs 7, 11) and the C and C' strands of the B7 isoforms¹² contribute to the binding interface. In addition, the structure identifies a number of amino acids involved in the binding interface, including Thr 53 and Tyr 105 of CTLA-4 and residues on the G strand of B7, which have not been previously reported. Although the residues in the BC loop (CDR1) of CTLA-4 do not directly contact B7, CDR1 packs against CDR3 and may influence its detailed conformation. Such an indirect effect is consistent with mutagenesis data showing that alterations in CDR1 affect the affinity for the B7 isoforms¹³. The extensive sequence conservation present at the B7 binding sites of CTLA-4 and CD28 and the concordance of CTLA-4 and CD28 mutagenesis

studies^{11,14,15} suggest that CTLA-4 and CD28 form similar interfaces with the B7 isoforms.

The CTLA-4 dimer interface in the complex is formed by residues that are positioned carboxy terminal to the G strand and residues centred around the A' strand, burying a total accessible surface area of 723 Å² (Fig. 2a, b). This newly discovered mode of dimerization results in a molecule that is substantially more extended than observed in other V-type dimers such as the CD8 homodimer¹⁶ and the VαVβ and VδVδ antigen receptor heterodimers^{17,18}, which form an extensive interface between their front sheets. Validation of this extended dimer is provided by the independent crystal structure of the unliganded, disulphide-linked human CTLA-4 dimer (J.-C.D.S. and X.Z., unpublished results), which exhibits the same dimer interface. Of note, comparison of the unliganded and liganded structures demonstrates that B7-2 binding does not induce any significant conformational reorganization in either the CTLA-4 monomer or the dimer interface, suggesting that unliganded CTLA-4 is poised in a 'pre-formed' productive binding mode. The organization of the human disulphide-linked CTLA-4 dimer places the B7-binding sites distal to the dimer interface and would be expected to support binding of two independent B7 molecules.

The residues that contribute to the human CTLA-4 dimer interface (Val 10, Leu 12, Ala 13, Ser 15, Tyr 115, Val 116, Ile 117, Asp 118, Pro 119 and Glu 120), as well as the disulphide linkage at Cys122 are nearly invariant in all known CTLA-4 sequences, which further supports the biological relevance of this interface (Fig. 2b, c). Although there is only modest sequence similarity between CTLA-4 and CD28 in this region, within the CD28 family there is considerable conservation, suggesting similar quaternary structures for both CTLA-4 and CD28 homodimers. Sequence divergence at this interface could provide a mechanism to prevent the formation

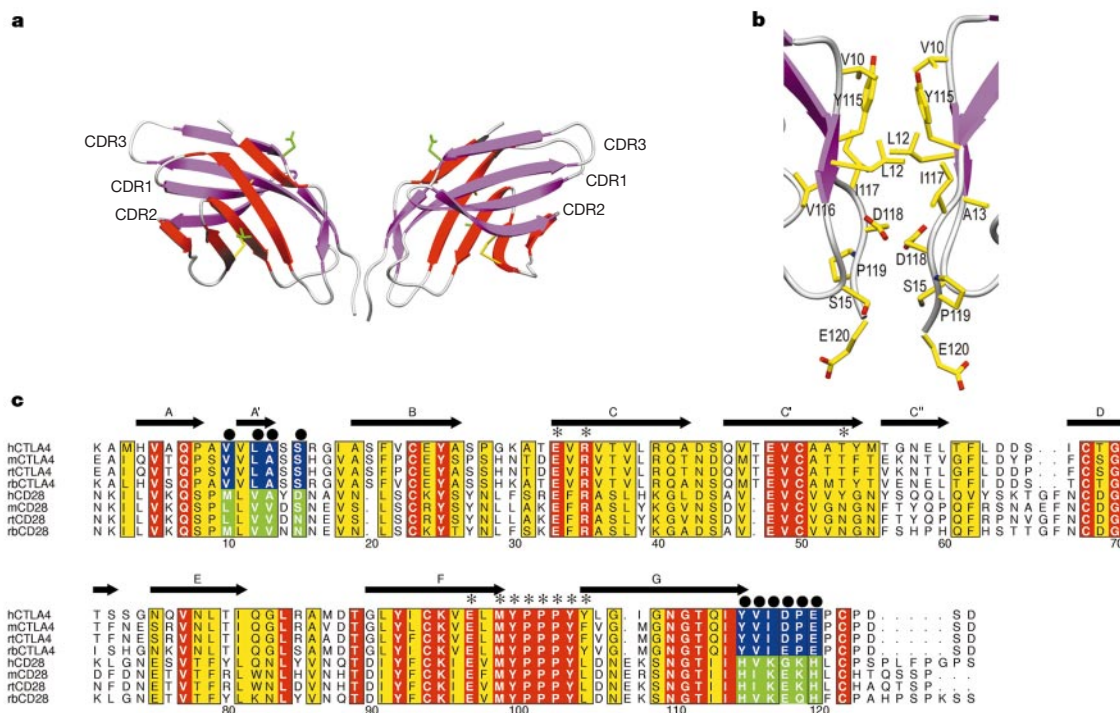


Figure 2 The disulphide-linked human CTLA-4 dimer. **a**, Ribbon diagram of the CTLA-4 dimer. The interface has contributions from residues C-terminal to the G strand and centred around the A' strand. The CDR analogous regions (CDR1, CDR2 and CDR3) are labelled and potential glycosylation sites are highlighted in green. The observed organization places the B7-binding sites distal to the dimer interface. **b**, Detailed view of the CTLA-4 dimer interface. This same interface is also present in the unliganded human CTLA-4 disulphide-linked dimer (J.-C.D.S. and X.Z., unpublished results). **c**, Sequence

alignment of the CTLA-4/CD28 family. Residues that participate in the human CTLA-4 dimer interface (filled circle) and B7-2-binding site (asterisk) are indicated. Residues with greater than 50% conservation are coloured in yellow, whereas those that are invariant are red. The residues in the human CTLA-4 dimer interface are coloured in blue, whereas the analogous residues in CD28 are green. H, human; m, murine; rt, rat; rb, rabbit. Strand colours are the same as in Fig. 1.

of CTLA-4/CD28 heterodimers and is consistent with the failure to detect such heterodimers *in vivo*¹⁹.

The B7-2 dimer interface is predominantly formed by the B, E and D strands and the BC, C'D and DE loops from the back sheet of each monomer (Fig. 3a, b) and buries a total accessible surface area of 1,405 Å², compared with 1,220 Å² in the B7-1 dimer interface⁸. This back-to-back packing arrangement is unusual for V-type dimers and has only been observed in the murine CTLA-4 (ref. 6) and human B7-1 non-covalent dimers⁸. Like CTLA-4, the B7-2 dimer observed in the crystal places the ligand-binding sites distal to the dimer interface, providing the potential for each B7-2 dimer to bind two CTLA-4 molecules. As with CTLA-4 and CD28, sequence divergence at the dimer interface may operate to prevent B7-1/B7-2 heterodimer formation (Fig. 3c).

The bivalent potential of both CTLA-4 and B7-2 is realized by the formation of a periodic assembly of dimeric molecules, characterized by a 103-Å repeat, that extends throughout the entirety of the crystal (Fig. 4). The direct observation of this periodic network immediately provides a model describing the assembly of these molecules at the T-cell/APC interface, and naturally suggests mechanisms that regulate the compartmentalization of signalling molecules at the immunological synapse. Such a network may promote the organized recruitment of inhibitory signalling molecules through interaction with either cytoplasmic or extracellular domains in CTLA-4 or B7, and may also decrease the local concentration of excitatory signalling molecules through simple steric crowding. The formation of an extended network may also act as a physical barrier to free diffusion that prevents entry and exit of molecules from the T-cell/APC interface. Given the similarities in their B7 recognition sites and modes of dimerization, some aspects of the organized assembly observed for CTLA-4 may also be operative in CD28-associated signalling.

Both B7 isoforms are composed of a membrane-distal receptor-binding domain and a membrane-proximal constant immuno-

globulin domain. By modelling the complete B7 molecule into the observed crystallographic repeat, the maximum dimension of the CTLA-4/B7-2 complex is estimated to span about 100 Å, excluding the residues in B7-2 and CTLA-4 that connect the immunoglobulin domains to the membrane-spanning segments

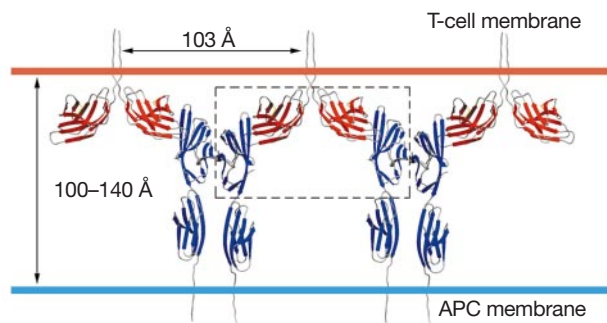


Figure 4 Molecular associations seen in the crystalline CTLA-4/B7-2 complex. The molecules in the dashed box represent the experimentally observed organization of CTLA-4 and the B7-2 receptor-binding domains within the crystal. The application of translational symmetry results in an alternating arrangement of bivalent dimers of CTLA-4 and B7-2 receptor-binding domains that extends throughout the entirety of the crystal. To model the interaction between CTLA-4 and the full-length B7 molecule, the membrane-proximal immunoglobulin domain from the human CD2 crystal structure was positioned on the basis of the known B7-1 crystal structure⁸. On the basis of the observed crystallographic repeat, this model predicts a characteristic spacing of 103 Å between adjacent CTLA-4 molecules and further suggests that the CTLA-4/B7-2 complex will impose a 100–140 Å spacing between the T cell and APC. The transmembrane and cytoplasmic regions are modelled to suggest that the periodic constraints of the extracellular domains are imposed on the cytoplasmic tails. Notably, all of the interfaces observed in the crystal are free of potential glycosylation sites that are present in CTLA-4 and B7-2. This model of association may extend to other members of this family, including CD28 and B7-1, as it is predicted that glycosylation would not preclude the formation of similar interfaces.

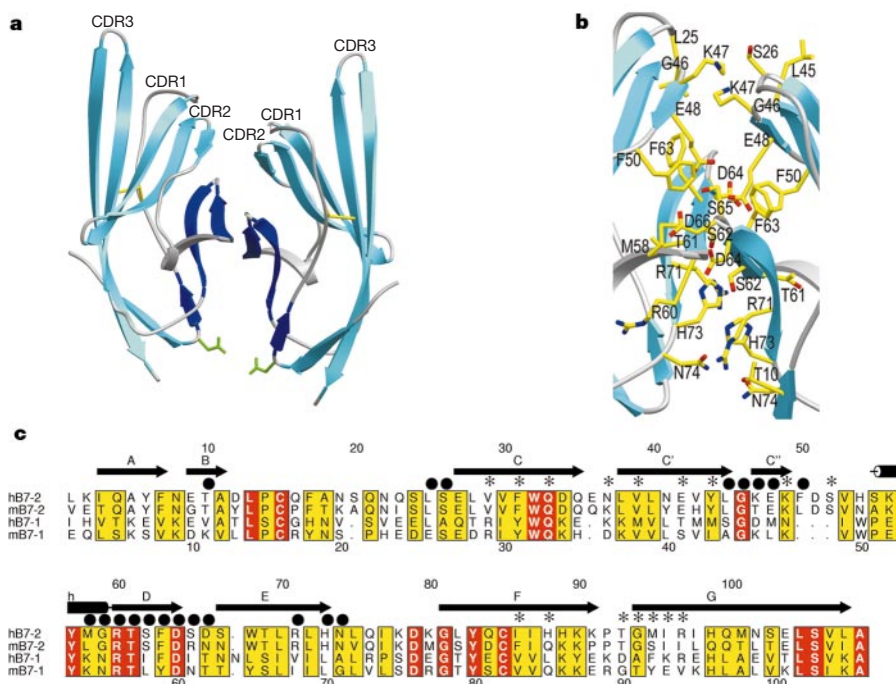


Figure 3 The human B7-2 dimer. **a**, Ribbon diagram of the B7-2 dimer. The interface is formed by the B, E, D, C' and C'' strands, and the BC, C'D and DE loops. The CDR analogous regions (CDR1, CDR2 and CDR3) are labelled and the potential glycosylation site is highlighted in green. This organization places the CTLA-4-binding sites distal to the dimer interface. The colour scheme is the same as in Fig. 1b. **b**, Detailed view of the B7-2 dimer interface. A similar interface is present in the unliganded human B7-1 structure¹¹

(See Methods). **c**, Structure-based sequence alignment of the B7-1 and B7-2 variable-like domains. Residues that participate in the human B7-2 dimer interface (filled circle) and CTLA-4-binding site (asterisk) are indicated. The numbering scheme for B7-2 and B7-1 are shown above and below the sequence alignment, respectively. Residues with greater than 50% conservation are coloured in yellow, whereas those that are invariant are red. Sequence abbreviations as in Fig. 2c.

(Fig. 4). This distance is consistent with the maximum dimensions of other receptor/ligand pairs that co-localize to the central region of the immunological synapse, including the MHC/TCR, CD2/CD58 and CD28/B7 complexes²⁰. In contrast, the length of the CTLA-4/B7-2 complex is significantly smaller than that of molecular species, such as the LFA-1/ICAM-1 complex, that localize in the peripheral zone of the synapse²⁰. Discrimination on the basis of molecular dimension may provide an efficient sorting mechanism for the compartmentalization observed at the immunological synapse, and is consistent with the disruption of the CD28/B7 complex through direct competition by CTLA-4.

We previously reported the structure of murine CTLA-4 (ref. 6) that formed a compact non-covalent dimer by means of packing of the back sheets, similar to that observed for the variable domains of B7-1 and B7-2. The C α of the last, firmly anchored residue in each murine monomer is about 22 Å from its symmetry mate, a distance that could reasonably be spanned by the remaining five residues in each chain to accommodate the formation of the required disulphide bond. Geometric considerations, however, indicate that the organization of the murine dimer is incompatible with the formation of a CTLA-4/B7 assembly at the T-cell/APC interface, as imposition of the observed human CTLA-4/B7-2 interface on the murine dimer results in a near-orthogonal disposition of the CTLA-4 and B7 dimer axes, an arrangement that is not consistent with known membrane topologies. Thus, the structure reported for the non-covalent murine dimer would seem not to have an obvious function in CTLA-4 signalling.

The structure of the human CTLA-4/B7-2 complex provides insights into the signalling mechanisms used by dimeric cell-surface receptors. One classical mechanism is used by the human growth hormone²¹ and fibroblast growth factor²² receptors that transduce signals only on ligand-induced dimerization. Another mechanism is used by the erythropoietin²³ and insulin²⁴ receptors, in which preformed receptor dimers undergo a significant conformational rearrangement on ligand binding to trigger signalling. CTLA-4 is an obligate dimer owing to the interchain disulphide linkage, and the present work shows that no conformational changes or reorganization of the dimer interface occur on binding of B7. These observations indicate that CTLA-4 is poised in a pre-formed productive binding mode before ligand engagement, and suggest that signalling relies on the assembly of at least two CTLA-4 dimers, or perhaps an extended network of multiple CTLA-4 and B7 dimers as observed in the crystal. This is in contrast to the erythropoietin, insulin and growth factor receptors^{21–24}, which are competent to signal as isolated dimeric molecules. Thus, the combination of both a pre-formed productive binding mode for CTLA-4 and the unique form of ordered clustering characteristic of the CTLA-4/B7 ligand pair may represent a distinct signalling mechanism available to dimeric cell-surface receptors. □

Methods

Crystallization

The soluble extracellular domain of human CTLA-4 (residues 1–126) was expressed in *Escherichia coli*, purified and refolded from inclusion bodies to yield the dimeric species covalently linked by a disulphide bond between Cys 122 of each monomer. We confirmed the presence of the inter-chain disulphide by reducing and non-reducing SDS–PAGE of dissolved crystals. The CTLA-4-binding domain of human B7-2 (residues 1–109) was similarly purified and refolded. The covalent CTLA-4 dimer and B7-2 domain were mixed in a 1:1 stoichiometry (based on monomer) at a total protein concentration of 3 mg ml^{−1}. Crystals of the complex were grown by hanging-drop vapour diffusion by mixing 2 µl each of protein solution and crystallization buffer composed of 18% PEG 20000, 100 mM HEPES, pH 7.0, followed by equilibration against the same buffer.

Data collection and structure determination

Diffraction from these crystals was consistent with the monoclinic space group *P*2₁ ($a = 47.85$, $b = 54.56$, $c = 103.09$ and $\beta = 91.63^\circ$), with two CTLA-4 monomers and two B7-2 monomers per asymmetric unit. Crystals were flash frozen at 100 K in mother liquor containing 15% glycerol. Data were collected at beamline X9B (National Synchrotron Light Source, Brookhaven National Laboratory) with 0.98 Å radiation, recorded on a 2 × 2 ADSC CCD detector and reduced with the HKL suite²⁵. The R_{merge} was 12.0% for all data

between 25–3.2 Å resolution (32.1% in the 3.3–3.2 Å shell) with overall completeness of 79% (79% in 3.3–3.2 Å shell). The average I/σ value is 6.6 for all data between 25–3.2 Å resolution (2.2 in the 3.3–3.2 Å shell). The structure was solved by molecular replacement, as implemented in CNS²⁶, using the 2.7-Å crystal structure of the B7-2 monomer (X.Z. and J.-C.D.S., unpublished results) and the murine CTLA-4 structure⁶ as search models.

Initially, the B7-2 monomer yielded two clear solutions related by approximate two-fold non-crystallographic symmetry, whereas the murine CTLA-4 monomer model yielded a single significant solution. Rigid-body refinement at 4.0 Å resolution resulted in the appearance of substantial density that was readily fit to a second CTLA-4 monomer by applying the non-crystallographic symmetry operator relating the two B7-2 monomers. After the removal of poorly ordered loops, rigid-body refinement at 4.0 Å resolution resulted in a R_{factor} and R_{free} of 37.2% and 40.7%, respectively. Modelling²⁷ and refinement at 3.2 Å resolution, including simulated annealing, bulk solvent correction and B-factor refinement²⁸, resulted in a final R_{factor} and R_{free} of 21.7% and 30.0%, respectively. The average B-factor for all atoms is 20.4 Å². The final model includes all residues except the two amino-terminal and five C-terminal residues in both CTLA-4 monomers, and residues 27–30 and 42–44 in one CTLA-4 monomer. The final model also has good geometry with 94.2% of residues falling in the most favourable and additionally allowed regions and only two residues (0.5%) in the disallowed region of the Ramachandran plot. r.m.s. bond lengths and r.m.s. bond angles are 0.008 Å and 1.3°, respectively, with all other stereochemical parameters comparable to or exceeding those of structures at similar resolution²⁸.

$R_{\text{merge}} = \sum |I_h - \langle I_h \rangle| / \sum I_h$, where $\langle I_h \rangle$ is the mean intensity for multiple measurements of symmetry-related reflection h , R_{factor} and $R_{\text{free}} = \sum_h ||F(h)_{\text{obs}}| - |F(h)_{\text{calc}}|| / \sum_h |F(h)_{\text{obs}}|$ for reflections in the working and test sets, respectively.

Structural analysis

Superposition of the two independent CTLA-4/B7-2 complexes in the asymmetric unit based on the B7-2 component revealed a 13° rotational difference between the two CTLA-4 molecules. The calculation was performed using the program FIT (G. Lu, <http://bioinfo1.mbfys.lu.se/~guoguang/fit.html>). The two independent CTLA-4 monomers that form the disulphide-linked dimer show close to a proper two-fold symmetry—being related by a rotation of 174.3° and a translation of 0.37 Å. The two monomers in the B7-2 dimer are not related by proper two-fold symmetry, as there is about a 175° rotation between the two monomers and about a 3.1 Å translation along the non-crystallographic symmetry axis, as calculated in FIT. The human B7-1 dimer shows perfect two-fold symmetry, as the molecular diad is coincident with a crystallographic two-fold axis⁸. Despite the deviation from perfect two-fold symmetry, most of the residues that contribute to the B7-2 interface occupy the same position in the primary sequence as those residues that form the B7-1 interface⁸.

The shape correlation statistic (Sc) measures the geometric complementarity at protein–protein interfaces, with Sc = 1 corresponding to a perfect fit and Sc = 0 corresponding to totally uncorrelated surfaces²⁹. Sc values in this study were calculated with the program SC²⁹ using default parameters. All buried surface areas were calculated with CNS²⁶ using a 1.4 Å probe. We prepared figures using the programs Setor³⁰ and ALSCRIPT³¹. We carried out structure-based sequence alignment of the B7 molecules with DALI³².

Received 3 January; accepted 16 February 2001.

- Oosterwegel, M. A., Greenwald, R. J., Mandelbrot, D. A., Lorsch, R. B. & Sharpe, A. H. CTLA-4 and T cell activation. *Curr. Opin. Immunol.* **11**, 294–300 (1999).
- Najafian, N. & Sayegh, M. H. CTLA4-Ig: a novel immunosuppressive agent. *Expert Opin. Invest. Drugs* **9**, 2147–2157 (2000).
- Garcia, K. C., Teyton, L. & Wilson, I. A. Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369–397 (1999).
- Masteller, E. L., Chuang, E., Mullen, A. C., Reiner, S. L. & Thompson, C. B. Structural analysis of CTLA-4 function *in vivo*. *J. Immunol.* **164**, 5319–5327 (2000).
- Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
- Ostrov, D. A., Shi, W., Schwartz, J. -C. D., Almo, S. C. & Nathenson, S. G. Structure of murine CTLA-4 and its role in modulating T cell responsiveness. *Science* **290**, 816–819 (2000).
- Metzler, W. J. *et al.* Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. *Nature Struct. Biol.* **4**, 527–531 (1997).
- Ikemizu, S. *et al.* Structure and dimerization of a soluble form of B7-1. *Immunity* **12**, 51–60 (2000).
- Wang, J. H. *et al.* Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**, 791–803 (1999).
- Shapiro, L. *et al.* Structural basis of cell–cell adhesion by cadherins. *Nature* **374**, 327–337 (1995).
- Morton, P. A. *et al.* Differential effects of CTLA-4 substitutions on the binding of human CD80 (B7-1) and CD86 (B7-2). *J. Immunol.* **156**, 1047–1054 (1996).
- Peach, R. J. *et al.* Both extracellular immunoglobulin-like domains of CD80 contain residues critical for binding T cell surface receptors CTLA-4 and CD28. *J. Biol. Chem.* **270**, 21181–21187 (1995).
- Peach, R. J. *et al.* Complementarity determining region 1 (CDR1)- and CDR3 analogous regions in CTLA-4 and CD28 determine the binding to B7-1. *J. Exp. Med.* **180**, 2049–2058 (1994).
- Kariv, I., Truneh, A. & Sweet, R. W. Analysis of the site of interaction of CD28 with its counter-receptors CD80 and CD86 and correlation with function. *J. Immunol.* **157**, 29–38 (1996).
- Truneh, A. *et al.* Differential recognition by CD28 of its cognate counter receptors CD80 (B7.1) and B70 (B7.2): analysis by site directed mutagenesis. *Mol. Immunol.* **33**, 321–334 (1996).
- Leahy, D. J., Axel, R. & Hendrickson, W. A. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 resolution. *Cell* **68**, 1145–1162 (1992).
- Garcia, K. C. *et al.* An alpha beta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**, 209–219 (1996).
- Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
- Lindsten, T. *et al.* Characterization of CTLA-4 structure and expression on human T cells. *J. Immunol.* **151**, 3489–3499 (1993).

20. Anton van der Merwe, P., Davis, S. J. & Dustin, M. L. Cytoskeletal polarization and redistribution of cell-surface molecules during T cell antigen recognition. *Semin. Immunol.* **12**, 5–21 (2000).
21. Cunningham, B. C. *et al.* Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. *Science* **254**, 821–825 (1991).
22. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).
23. Wilson, I. A. & Jolliffe, L. K. The structure, organization, activation and plasticity of the erythropoietin receptor. *Curr. Opin. Struct. Biol.* **9**, 696–704 (1999).
24. Luo, R. Z., Beniac, D. R., Fernandes, A., Yip, C. C. & Ottensmeyer, F. P. Quaternary structure of the insulin–insulin receptor complex. *Science* **285**, 1077–1080 (1999).
25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **176**, 307–326 (1997).
26. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
27. Jones, T. A., Cowan, S., Zou, J. Y. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **8**, 631–639 (1998).
29. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
30. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138 (1993).
31. Barton, G. J. Protein multiple sequence alignment and flexible pattern matching. *Methods Enzymol.* **183**, 403–428 (1990).
32. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).

Acknowledgements

We thank M. Scharff, P. Scherer, A. Davidson, A. Bresnick, T. DiLorenzo, A. Kalergis and M. Roden for comments. We also thank K. Rajashankar for assistance with data collection. This work was supported by grants from the National Institute of Allergies and Infectious Diseases. We acknowledge the support of the Albert Einstein Comprehensive Cancer Center.

Correspondence and requests for materials should be addressed to S.C.A. (e-mail: almo@aecom.yu.edu) or S.G.N. (e-mail: nathenso@aecom.yu.edu).

Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses

Carin C. Stamper^{*}, Yan Zhang^{†‡}, James F. Tobin^{*}, David V. Erbe^{*}, Shinji Ikemizu[‡], Simon J. Davis[§], Mark L. Stahl^{*}, Jasbir Seehra^{*}, William S. Somers^{*} & Lidia Mosyak^{*}

^{*} Departments of Biological Chemistry and Musculoskeletal Sciences, Wyeth Research, 87 Cambridge Park Drive, Cambridge, Massachusetts 02140, USA

[‡] Division of Structural Biology, The Henry Wellcome Building for Genomic Medicine, The University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

[§] Nuffield Department of Clinical Medicine, The University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

Optimal immune responses require both an antigen-specific and a co-stimulatory signal. The shared ligands B7-1 and B7-2 on antigen-presenting cells deliver the co-stimulatory signal through CD28 and CTLA-4 on T cells. Signalling through CD28 augments the T-cell response, whereas CTLA-4 signalling attenuates it. Numerous animal studies^{1,2} and recent clinical trials^{3,4} indicate that manipulating these interactions holds considerable promise for immunotherapy. With the consequences of these signals well established, and details of the downstream signalling events emerging^{5–7}, understanding the molecular nature of these extracellular interactions becomes crucial. Here we report the crystal structure of the human CTLA-4/B7-1 co-stimulatory complex at 3.0 Å resolution. In contrast to other interacting cell-surface

molecules, the relatively small CTLA-4/B7-1 binding interface exhibits an unusually high degree of shape complementarity. CTLA-4 forms homodimers through a newly defined interface of highly conserved residues. In the crystal lattice, CTLA-4 and B7-1 pack in a strikingly periodic arrangement in which bivalent CTLA-4 homodimers bridge bivalent B7-1 homodimers. This zipper-like oligomerization provides the structural basis for forming unusually stable signalling complexes at the T-cell surface, underscoring the importance of potent inhibitory signalling in human immune responses.

The extracellular domains of human soluble (s) CTLA-4, residues 3–125 (SWISSPROT numbering), and sB7-1, residues 1–214, were expressed in Chinese hamster ovary (CHO) cells. The complex containing fully glycosylated sCTLA-4 and sB7-1 was crystallized in the space group C222₁. In the crystal lattice, sCTLA-4 and sB7-1 monomers each associate as non-crystallographic, roughly two-fold symmetric homodimers; single copies of each homodimer together form the asymmetric unit (Fig. 1a). The association of the four monomers in the asymmetric unit is driven by the interaction of three surfaces: the surface mediating sB7-1 homodimerization; the CTLA-4 homodimer interface; and the receptor–ligand binding interface.

The structures of the uncomplexed sCTLA-4 (refs 8, 9) and sB7-1

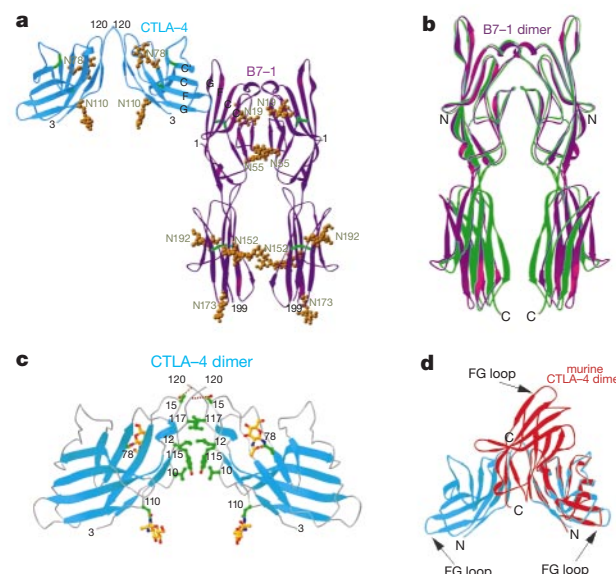


Figure 1 Structural comparison of sCTLA-4/sB7-1 complex with uncomplexed forms of sCTLA-4 and sB7-1. **a**, Ribbon diagram of the sCTLA-4/sB7-1 complex showing two sB7-1 (purple) and two sCTLA-4 (cyan) molecules in the asymmetric unit. Disulphide bonds (green) and sugar moieties (yellow) are also shown. β -sheets involved in the receptor–ligand interaction are labelled. Glycosylation sites on sCTLA-4 (Asn 78 and Asn 110) and on sB7-1 (Asn 19, Asn 55, Asn 152, Asn 173 and Asn 192) are all surface exposed. None of the ordered glycosides is involved in the receptor–ligand recognition. **b**, Superposition of the ligand-binding V-set domains of complexed (purple ribbons) and uncomplexed (cyan ribbons) forms of sB7-1. Apparent displacement of the membrane-proximal domains does not affect the interactions at the interdomain regions and is perhaps a result of differences in crystal packing contacts. **c**, Detailed view of the sCTLA-4 homodimer interface. Residues involved in hydrophobic interactions between the two monomers are shown in green. Two hydrogen bonds formed at the elbow of the homodimer are indicated as red dashed lines. Spatial proximity of the C termini (Glu 120) would enable a formation of disulphide bond between cysteines at positions 122. Asn 110 is the only strictly conserved glycosylation site in CTLA-4 and CD28. Located between the two monomers, N-glycans in these positions might stabilize the homodimer. **d**, Superposition of human sCTLA-4 dimer (cyan) on murine sCTLA-4 dimer (red). From this orientation, the contradiction between the two dimerization modes is evident. R.m.s. deviation between the two individual monomers is 1.5 Å for 114 C α pairs. The FG loop of sCTLA-4 in the complex is displaced by 2.5 Å toward sB7-1. Figure prepared with RIBBONS²⁹.

[†] Present address: Pfizer Cambridge Discovery Technology Center, 620 Memorial Drive, Cambridge, Massachusetts 02139, USA.

(ref. 10) have shown that the two proteins comprise, respectively, a single V-set, and paired V-set and C1-set immunoglobulin superfamily (IgSF) domains. Overall, the sCTLA-4 and sB7-1 monomers do not show any significant conformational rearrangements on complex formation, apart from small local changes restricted to their binding sites. The glycosylated sB7-1 homodimer seen in the crystals of the complex is essentially identical to that observed in crystals of deglycosylated sB7-1 (ref. 10): the monomer–monomer interface is the same size ($570 \text{ \AA}^2$ per monomer; calculated using the program SURFACE¹¹; probe radius $1.4 \text{ \AA}$) and involves the same residues. When the V-set domains of the two sB7-1 homodimers are superimposed (Fig. 1b), the membrane-proximal C1-set domains are displaced by only $3 \text{ \AA}$. Similar variation has been observed in individual crystals of sB7-1 (S.I. *et al.*, unpublished data), suggesting that these differences are more likely to result from variability in sB7-1 self-association than from the effects of ligand binding. sB7-1 homodimerization has been confirmed in solution using analytical ultracentrifugation methods¹⁰.

Although the structure of the sCTLA-4 monomer is very similar to that of the murine sCTLA-4 monomer⁸, the human and murine⁸ sCTLA-4 homodimers observed in the two lattices are very different (Fig. 1c, d). Of the two structures, we consider that the sCTLA-4 homodimer present in the crystals of the complex represents the more plausible model for native CTLA-4. First, all of the interacting residues are strictly conserved, in contrast to the conservation of only four of the ten residues forming the murine dimer interface⁸. Second, the human sCTLA-4 homodimer is not precluded by glycosylation. If used as in human CTLA-4 (ref. 12), an *N*-linked glycan at the conserved Asn 78 glycosylation site would prevent murine-like homodimerization in all species. Third, the human homodimer allows co-ligation of sB7-1 molecules around an axis orthogonal to the membrane, maintaining the $\sim 140 \text{ \AA}$ intermembrane distance thought to be a critical feature of the immunological synapse⁵. Last, the interchain disulphide bond will more readily form in the human sCTLA-4 homodimer. The $7.7 \text{ \AA}$ spacing between the $C\alpha$ atoms of the last visible residue in each monomer (Pro 119) is more likely to be bridged by the three additional residues and by the disulphide bond in the human sCTLA-4 dimer than is the equivalent $30.1 \text{ \AA}$ distance in the murine structure. Consequently, the murine model for CTLA-4 dimerization and signalling proposed previously⁸ is most probably not relevant to the formation of co-stimulatory complexes.

The human sCTLA-4 monomers interact through residues in the

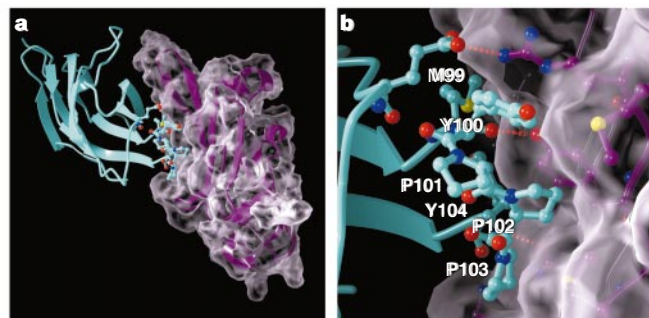


Figure 2 Overview of receptor–ligand interactions. **a**, Ribbon diagram showing orthogonal interaction between sCTLA-4 (cyan) and sB7-1 (purple) monomers. Also shown is the molecular surface representation (white transparent) of the ligand-binding domain of sB7-1 to emphasize the high geometric match between the two interacting surfaces. **b**, Direct receptor–ligand contacts. The ⁹⁹MYPPPY¹⁰⁴ loop of sCTLA-4 is buried in a shallow depression of the sB7-1 GFCC' surface. Colour coding is as in **a**. Three out of five hydrogen bonds formed across the β -sheets of the interacting domains are depicted as red dashed lines. Several other side chains on CTLA-4 and B7-1 (not shown) may contribute to the binding through appreciable, but not direct, contacts formed on the periphery of the binding interface. Figure prepared with BobScript³⁰.

A and G β -strands (Val 10, Leu 12, Ser 15, Tyr 115, Ile 117, Glu 120), burying only $460 \text{ \AA}^2$ of surface¹¹ per monomer. The interface is essentially hydrophobic except for reciprocal, interchain hydrogen bonds between the side chain of Ser 15 and the main-chain nitrogen of Glu 120 at the elbow of the homodimer (Fig. 1c). This weak interaction might explain why, in the absence of the interchain disulphide (Cys–Cys 122), sCTLA-4 is monomeric in solution (ref. 9; and Y.Z. and M.S., unpublished data). The amino acids mediating CTLA-4 dimerization are not conserved in CD28; however, the equivalent residues exhibit the same overall hydrophobicity, implying that CD28 might form similar homodimers.

Differences noted⁸ between the murine structure and the human sCTLA-4 structure determined by NMR methods⁹ are also seen in the complex crystals. In particular, the V-set domain has C'DEBA:GFCC' rather than DEBA:GFCC'C' topology, and the three consecutive prolines in the FG loop, MY¹⁰¹PPP¹⁰³Y, adopt the *cis*–*trans*–*cis* rather than *trans*–*trans*–*cis* conformation seen in the NMR structure⁹. This has important implications for B7-1 binding (see below).

Mutational data have implicated the membrane-proximal domain of B7-1 in interactions with CTLA-4 and CD28 (ref. 13), but these effects must have been indirect as sCTLA-4 only contacts residues of the sB7-1 V-set domain. Receptor–ligand binding occurs through the GFCC' face of the sCTLA-4 and sB7-1 V-set domains, with a roughly 90° angle between the interacting β -sheets (Fig. 2a). In this orientation, the FG, BC and C'C' loops of sCTLA-4 extend across the base of the sB7-1 five-stranded β -sheet, burying a total of $1,255 \text{ \AA}^2$ of solvent-accessible surface¹¹ ($600 \text{ \AA}^2$ from sB7-1 and $655 \text{ \AA}^2$ from sCTLA-4). These values are at the low end of the range for protein–protein binding sites (600 – $900 \text{ \AA}^2$)¹⁴.

Orthogonal binding mediated by the GFCC'C' face of the V-set IgSF domains of cell-surface molecules was first seen in rat and human sCD2 crystal contacts^{15,16}, and subsequently in the native complex of sCD2 and sLFA-3 (ref. 17) and in crystals of the coxsackie virus and adenovirus receptor¹⁸. It seems likely that this orthogonal binding mode will be a recurrent theme of V-set interactions. There is no compelling reason why such orthogonal interactions should have arisen through convergent evolution, but it is possible that these complexes bear the imprint of primordial IgSF interactions.

Interatomic contacts (85 total; van der Waals radius $3.9 \text{ \AA}$) are made between 13 residues of CTLA-4 residues and 13 residues of

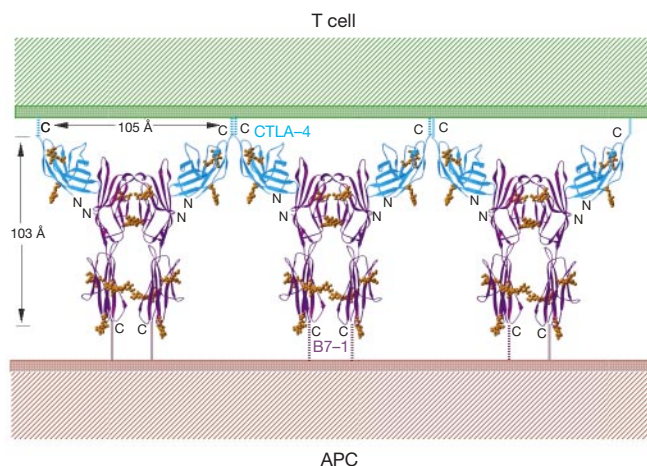


Figure 3 Molecular association of sCTLA-4 and sB7-1 in the crystal lattice. Shown are 'skewed zipper' arrays in which sCTLA-4/sB7-1 complexes would be evenly spaced along membrane surfaces with a separation of $105 \text{ \AA}$. In the perpendicular direction, across membranes, ligated receptors would span $140 \text{ \AA}$. Geometrically, sugar chains attached at Asn 173 on B7-1 (bottom) are close to the cell membrane, implying their potential involvement in interaction with the membrane, perhaps by stabilizing the orientation of the B7-1 dimers. APC, antigen-presenting cell. Figure prepared with RIBBONS²⁹.

B7-1. Most of these are hydrophobic contacts, but there are also five hydrogen bonds. The FG loop of sCTLA-4, which contains the hydrophobic ⁹⁹MYPPPY¹⁰⁵ sequence that is strictly conserved in CTLA-4 and CD28, dominates the interaction and contributes 400 Å² of protein surface¹¹ to the binding interface (Fig. 2b). The FG loop makes hydrophobic contacts with a largely nonpolar surface of sB7-1 consisting of Tyr 31, Met 38, Thr 41, Met 43, Val 83, Leu 85, Ala 91, Phe 92 and Leu 97.

Mutations of the two central residues, Tyr 31 and Met 38, disrupt B7-1 binding to CTLA-4 and CD28 (ref. 13); the other residues forming this surface have not been tested. Five hydrogen bonds, including one between the charged residues Glu 33 on sCTLA-4 and Arg 29 on sB7-1 are likely to contribute to the specificity of binding (Fig. 2b). All residues in the ⁹⁹MYPPPY¹⁰⁵ sequence, except Pro 101, are in direct contact with sB7-1, and all alanine substitutions of these residues, including Pro 101, reduce or abolish binding to B7-1 (ref. 9). Analysis of the complex structure indicates that the crucial role of Pro 101 is to initiate the *cis-trans-cis* main-chain conformation of the FG loop, directing Tyr 100, Pro 102, Pro 103 and Tyr 104 towards interactions with sB7-1. At the core of the interface, Pro 102 of CTLA-4 and Tyr 31 of B7-1 participate in a stacking interaction (Figs 2b, 4) that is likely to be conserved both across species and in interactions involving CD28 and B7-2 (in B7-2 Tyr 31 is replaced by phenylalanine).

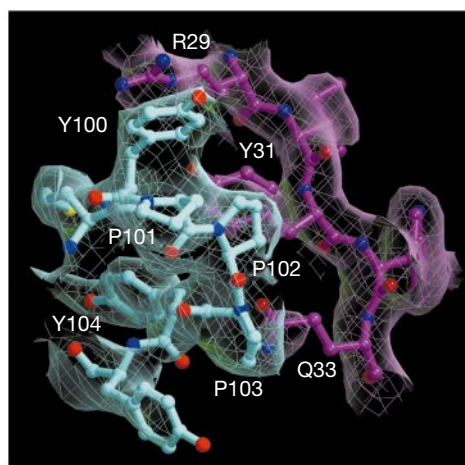


Figure 4 A 3 Å resolution electron-density map in the region of the receptor–ligand binding site. Electron density is from an annealed omit map calculated using the $3F_{\text{obs}} - 2F_{\text{calc}}$ amplitudes and model phases, with the CTLA-4 atoms excluded from the refinement and all calculations. The map is contoured at 1σ level. Colour coding is as in Figs 1 and 2.

Table 1 Statistics for data collection and refinement

Data collection		
Resolution range (Å)	20–3.0	3.11–3.0
Completeness (%)	99.2	94.1
Total observations	209,381	—
Unique reflections	38,815	3,636
Average $I/\sigma(I)$	19.1	3.0
$R_{\text{sym}}^*(\%)$	6.7	35.1
Model refinement		
Maximum resolution (Å)	3.0	—
Number of reflections (free)	$F_{\text{obs}} > 2\sigma$ 31,410 (1,887)	$F_{\text{obs}} > 0\sigma$ 35,465
$R_{\text{work}}/R_{\text{free}}^{\dagger}(\%)$	23.1/25.9	25.3/28.2
r.m.s. deviations		
Bonds (Å)	0.010	
Angles (°)	1.48	

* $R_{\text{sym}} = \sum |I_h - \langle I_h \rangle| / \sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalents. Numbers in the second column reflect statistics for the last resolution shell.

$\dagger R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, R_{free} is equivalent to R_{work} , but calculated for a randomly chosen 6% of reflections omitted from the refinement process.

Accordingly, the two binding surfaces exhibit a very high degree of shape complementarity. An algorithm¹⁹ measuring the degree of geometric fit between two protein surfaces gives scores of 0.74–0.77 for the sCTLA-4/sB7-1 binding interface. This value is similar to those for constitutive oligomeric proteins (0.7–0.76) and much higher than scores for protein antigen–antibody interfaces (0.64–0.68) or interacting cell-surface molecules (0.45–0.58; ref. 17 and references therein). This degree of complementarity accounts for the enthalpy-driven nature of binding (R. O'Brien *et al.*, personal communication), whereas the small interacting surfaces and slightly unfavourable entropy explain the very fast binding kinetics¹².

In the crystal lattice, the sCTLA-4 and sB7-1 homodimers pack together to form a periodic arrangement in which bivalent sCTLA-4 homodimers bridge bivalent sB7-1 homodimers (Fig. 3). The sB7-1 and shorter CTLA-4 homodimers associate orthogonally, thus generating a 'skewed zipper' arrangement. It has long been clear that CTLA-4 exists as a constitutive, bivalent homodimer, and the affinity of sB7-1 self-association (dissociation constant, $K_d = 20$ – $50 \mu\text{M}$) indicates that B7-1 is also likely to exist as a dimer at the cell surface, albeit in dynamic monomer–dimer equilibrium¹⁰. We therefore expect oligomeric arrays that are similar, if not identical, to those seen in the crystals to form at the membrane interface between T cells and antigen-presenting cells.

Including the 'stalk' regions, the extracellular domains of the ligated receptors are expected to span a distance of $\sim 140 \text{ Å}$ between the opposing cell membranes, a distance compatible with that required by T-cell receptor (TCR)–MHC, natural killer cell inhibitory receptor (KIR)–MHC or CD2–CD58 intercellular interactions^{5,17,20,21}. Overall, this arrangement is reminiscent of the 'cell-adhesion zipper' observed in the crystals of cadherins, which is thought to be a fundamental feature in adhesive interactions between cells expressing these molecules²².

As far as is generally known, cell-surface molecules bind their ligands monovalently and with very low (micromolar) affinities²³. The submicromolar affinity of B7-1 for CTLA-4 ($K_d = 0.2$ – $0.4 \mu\text{M}$; ref. 12) is thus unusually high for interacting cell-surface molecules. Our structural analysis of the complex implies that potent B7-mediated inhibitory signalling is not based exclusively on the stability of association between individual homodimers, but rather that the counter-receptor oligomeric arrays will also strengthen the interaction between the opposing cells. The combination of sub-micromolar affinity and oligomeric, high-avidity binding is, to our knowledge, unique to CTLA-4/B7-1 interactions.

The amplification and attenuation of TCR signalling occurs in a specialized contact area on the cell surface, termed the immunological synapse^{5,6}, through the recruitment of receptors and their signalling complexes that effect the phosphorylation state of the TCR (ref. 7). Notably, inhibition of TCR signalling by CTLA-4 has been seen only to occur when both signals are delivered by the same cell surface²⁴; however, other data suggest that the regulatory signals delivered by CD28 and CTLA-4 may not be strictly cell autonomous²⁵. Regardless of the precise nature of the signals and mechanisms involved, it now seems that the inhibition of immune responses by CTLA-4 requires the formation of unusually stable complexes, the structure of which is revealed by the crystals of sCTLA-4/sB7-1. Whether such stable complexes are required for the efficient use of low-abundance CTLA-4 molecules at the T-cell surface, or enhance the potency of inhibitory CTLA-4 signalling by concentrating key effector molecules such as the phosphatase SHP-2 (ref. 7) under the immunological synapse, or both, remains to be established. □

Methods

Production and purification of B7-1 and CTLA-4

Constructs encoding the extracellular portions of human CTLA-4 or B7-1 were fused to an enterokinase cleavage sequence (DYKDDDDK) followed by the IgG1 CH2–CH3 domains, and expressed in CHO cells. In addition, the cysteine at position 122 of CTLA-4

was mutated to serine. These were recovered from conditioned medium by protein-A chromatography, cleaved with enterokinase, further purified and confirmed to be monomeric by gel-filtration chromatography as described²⁶. Appropriate binding characteristics for each protein were confirmed by surface plasmon resonance (Biacore) analysis and found to be consistent with published values¹². For CTLA-4, the final cleaved protein comprised residues 3–125 (with the C122S mutation) followed by the octa-peptide DYKDDDDK at the carboxy terminus. For B7-1, the cleaved protein consisted of residues 1–214 with the DYKDDDDK octa-peptide again forming the C terminus.

Crystallization and data collection

The B7-1/CTLA4 complex was formed by incubating a 1:1.5 molar ratio of B7-1 and CTLA4 and purified using a TSK-3000SW size exclusion column (TosoHaas). The purified complex was buffer exchanged and concentrated to 10 mg ml⁻¹ in 20 mM Tris-HCl, pH 8.0. Crystals were grown at 18 °C by the hanging drop method by combining 2 µl of protein solution with 2 µl of 14% PEG 8000, 200 mM magnesium acetate and 100 mM cacodylate, pH 6.3, and equilibrated against 1 ml of the same solution. Crystals appeared in 5 days and were soaked in 16% PEG 8000, 200 mM magnesium acetate, 20 mM Tris-HCl pH 8.0, 100 mM cacodylate, pH 6.3, 20% ethylene glycol for about 1 min before plunging into liquid nitrogen. Crystals belong to space group C222₁ with unit-cell dimensions $a = 88.5$, $b = 183.4$, $c = 230.8$ Å and a calculated solvent content of 58% for two B7-1 and two CTLA-4 molecules per asymmetric unit (adding 82,000 (82K) of glycosylation estimated from SDS-PAGE gel mobility). The 3.0 Å resolution data were collected from a single crystal at 100K at Beamline 5.0.2 at the Advanced Light Source using a Quantum 4 CCD. 725 0.2° rotation images were collected and reduced with HKL2000 (ref. 27) giving statistics outlined in Table 1.

Structure determination and refinement

The structure was solved by molecular replacement in CNS²⁸ using the crystallographic dimer observed in the human B7-1 structure as the search model¹⁰. After rigid body refinement of each B7-1 domain the R factor was 55.3%. Initial electron-density maps phased with B7-1 alone showed clear electron density for the two CTLA-4 molecules. The model of murine CTLA-4 monomer (PDB accession number 1dqt) was placed into electron density and was manually rebuilt in QUANTA (Molecular Simulations) to reflect the sequence of human CTLA-4. Non-conserved amino-acid residues and loop regions of CTLA-4 and B7-1, as well as sugar moieties, were entirely built into electron density. A composite annealed omit map was calculated to check the conformation of CTLA-4 and B7-1 loop regions. All refinement procedures were done in CNS and used data from 20.0–3.0 Å. A 2σ cut-off was applied during the refinement because of the anisotropy of the data. Six per cent of the reflections were randomly selected for generation of the R_{free} data set (see Table 1). The anisotropy of the data was corrected for also using CNS.

Tight non-crystallographic symmetry restraints were gradually relaxed during the refinement process resulting in a final model with R_{work} of 23.2% and R_{free} of 25.9%. Continuous electron density was seen for sugar moieties at four N-linked carbohydrate sites on CTLA-4 dimer (Asn 78 and Asn 110) and at ten glycosylation sites on B7-1 dimer (Asn 19, Asn 55, Asn 152, Asn 173 and Asn 192). No saccharides could be modelled at six other expected glycosylation sites on B7-1 (Asn 64, Asn 177 and Asn 198). The final model contains two B7-1 molecules (residues A1–A199 and B1–B199), two CTLA-4 molecules (residues C3–C120 and D3–D120), 20 monosaccharides and shows reasonable geometry: 98.9% of non-glycine ϕ and ψ angles lie in the allowed regions of the Ramachandran plot (77.6% in the most favoured regions and 21.4% in additionally allowed regions); six amino-acid residues (1.1%) have disallowed angles.

Received 27 December 2000; accepted 9 February 2001.

- Lenschow, D. J., Walunas, T. L. & Bluestone, J. A. CD28/B7 system of T cell costimulation. *Annu. Rev. Immunol.* **14**, 233–258 (1996).
- Greenfield, E. A., Nguyen, K. A. & Kuchroo, V. K. CD28/B7 costimulation: a review. *Crit. Rev. Immunol.* **18**, 389–418 (1999).
- Guinan, E. C. *et al.* Transplantation of anergic histoincompatible bone marrow allografts. *N. Engl. J. Med.* **340**, 1704–1714 (1999).
- Abrams, J. R. *et al.* Blockade of T lymphocyte costimulation with cytotoxic T lymphocyte-associated antigen 4-immunoglobulin (CTLA4Ig) reverses the cellular pathology of psoriatic plaques, including the activation of keratinocytes, dendritic cells, and endothelial cells. *J. Exp. Med.* **192**, 681–694 (2000).
- van der Merwe, P. A., Davis, S. J., Shaw, A. S. & Dustin, M. L. Cytoskeletal polarization and redistribution of cell-surface molecules during T cell antigen recognition. *Semin. Immunol.* **12**, 5–21 (2000).
- Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
- Lee, K.-M. *et al.* Molecular basis of T cell inactivation by CTLA-4. *Science* **282**, 2263–2266 (1998).
- Ostrov, D. A., Shi, W., Schwartz, J.-C. D., Almo, S. C. & Nathenson, S. G. Structure of murine CTLA4 and its role in modulating T cell responsiveness. *Science* **290**, 816–819 (2000).
- Metzler, W. J. *et al.* Solution structure of human CTLA4 and delineation of a CD80/CD86 binding site conserved in CD28. *Nature Struct. Biol.* **4**, 527–531 (1997).
- Ikemizu, S. *et al.* Structure and dimerization of a soluble form of B7-1. *Immunity* **12**, 51–60 (2000).
- Collaborative computational project No. 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *J. Exp. Med.* **185**, 393–403 (1997).
- Peach, R. J. *et al.* Both extracellular immunoglobulin-like domains of CD80 contain residues critical for binding T cell surface receptors CTLA-4 and CD28. *J. Biol. Chem.* **268**, 21181–21187 (1993).
- Janin, J. & Chothia, C. The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030 (1990).
- Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* **360**, 232–239 (1992).
- Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I. & Davis, S. J. Crystal structure of the extracellular

- region of the human cell adhesion molecule CD2 at 2.5 Å resolution. *Structure* **15**, 755–766 (1994).
- Wang, J.-h. *et al.* Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **7**, 791–803 (1999).
- van Raaij, M. J., Chouin, E., van der Zandt, H., Bergelson, J. M. & Cusack, S. Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 Å resolution. *Struct. Fold Des.* **8**, 1147–1155 (2000).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Garboczi, D. N. *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134–141 (1996).
- Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G. & Sun, P. D. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* **405**, 537–543 (2000).
- Shapiro, L. *et al.* Structural basis of cell–cell adhesion by cadherins. *Nature* **374**, 327–337 (1995).
- Davis, S. J., Ikemizu, S., Wild, M. K. & van der Merwe, P. A. CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunol. Rev.* **163**, 217–236 (1998).
- Griffin, M. D. *et al.* Blockade of T cell activation using a surface-linked single-chain antibody to CTLA-4 (CD152). *J. Immunol.* **164**, 4433–4442 (2000).
- Bachmann, M. F., Kohler, G., Ecabert, B., Mak, T. W. & Kopf, M. Lymphoproliferative disease in the absence of CTLA-4 is not T cell autonomous. *J. Immunol.* **163**, 1128–1131 (1999).
- Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to sLex and PSGL-1. *Cell* **103**, 467–479 (2000).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brunger, A. T. *et al.* Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).
- Esnouf, R. M. An extensively modified version of Molscript that includes enhanced colouring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).

Acknowledgements

We thank R. Zollner and his colleagues for large scale cell culture, and D. I. Stuart and E. Y. Jones for helpful comments. We also thank the staff at Advanced Light Source for assistance with data collection. S.I. and S.J.D. are supported by the Wellcome Trust and the Human Frontier Science Programme.

Correspondence and requests for materials should be addressed to W.S.S. (e-mail: wsomers@genetics.com) or L.M. (e-mail: lmosyak@genetics.com). The coordinates have been deposited in the Protein Data Bank. The accession code is 1I84.

corrections

Genetic control and evolution of sexually dimorphic characters

Artyom Kopp, Ian Duncan, Dorothea Godt & Sean B. Carroll

Nature **408**, 553–559 (2000).

The name of one author, Dorothea Godt, was omitted from the author list but is included above. She is affiliated with the Department of Zoology, University of Toronto, Canada. In addition, a gene name was misspelled: the correct spelling is *bric à brac*. □

Mitochondrial genome variation and the origin of modern humans

Max Ingman, Henrik Kaessmann, Svante Pääbo & Ulf Gyllensten

Nature **408**, 708–712 (2000).

The complete mtDNA sequences upon which the analyses are based can be found at our website (<http://www.genpat.uu.se/mtDB/>), dedicated to the analysis of complete mtDNA genomes of humans. The 53 sequences are also available in Genbank with the accession numbers AF346963–AF347015. □

Are you getting enough protein?

New ideas in proteomics and protein handling.

oMALDI source

Applied Biosystems www.appliedbiosystems.com

A new angle on proteomics

Applied Biosystems' new orthogonal matrix-assisted laser desorption/ionization (oMALDI) source for the company's API QSTAR Pulsar Hybrid LC/MS/MS system features interchangeable MALDI and API ion sources and is compatible with tandem MS. With an oMALDI source, the QSTAR Pulsar system identifies proteins separated by one- or two-dimensional gel electrophoresis at the femtomole level by peptide mass fingerprinting of the protein digest or using sequence data from tandem mass spectra acquired by collision-induced dissociation of individual peptide precursor ions. Database search and acquisition of MS and MS/MS spectra can be combined interactively for more informative analytical data.

Reader Service No. 100

FluoroTect Green_{lys}

Promega www.promega.com

In vitro translation labelling system

Based on a lysine-charged tRNA labelled at the epsilon position of the lysine with the fluorophore BODIPY, FluoroTect Green_{lys} allows for the fluorescent labelling and detection of proteins synthesized *in vitro*. Fluorescent lysine residues are incorporated into synthesized proteins during *in vitro* translation reactions, eliminating the need for radioactivity. There is no need to transfer, fix or dry gels.

Reader Service No. 101

ADV01

Cytoskeleton, Inc. www.cytoskeleton.com

Bone up on protein measurement

This advanced protein assay reagent combines low protein-to-protein variance with a strong signal for a sensitive assay. A one-step procedure results in a change from green to blue recognized by measuring absorbance at 570–615 nm. ADV01 can be used to determine protein concentration in a buffer, biological fluid, tissue culture media or in combination with an activity assay to follow protein purification.

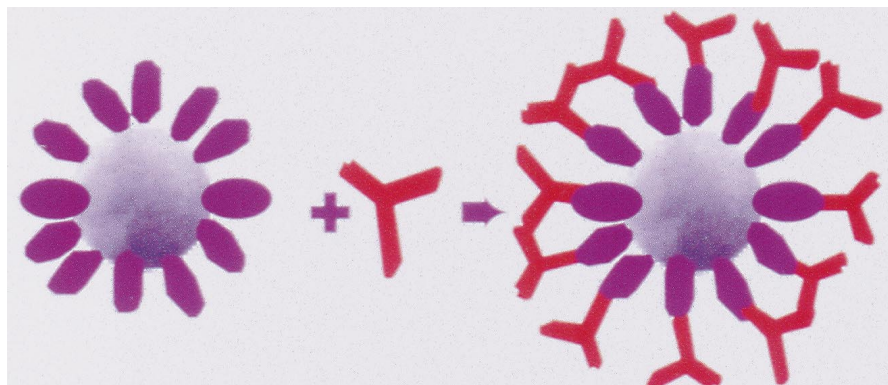
Reader Service No. 102

Magnetic protein A particles

CPG www.cpg-biotech.com

Immunoassay and bioassay supports

Protein A covalently attached to magnetic porous glass (MPG) particles confers immunoglobulin binding properties similar to those of unconjugated protein A. Nominal binding capacity of 1 mg of particles is 40–75 µg



More MPG: protein A gives solid support.

immunoglobulin. MPG protein A particles can be used as solid supports in single or automated immunoassays and bioassays for separating unbound labelled tracer. They are suitable for small-scale affinity purifications or immunoprecipitation direct from ascites, serum, homogenates, lysates and culture supernatants for gel or other physical analyses. The particles can be regenerated up to 12 times and will continue to provide more than 90% of initial binding capacity.

Reader Service No. 103

ArrayStat

Imaging Research www.imagingresearch.com

Inferential software

This genomics software package is designed to provide a comprehensive inferential statistical analysis of gene expression array data from a variety of file formats with as few as two replicates per spot. It can be used to identify statistically significant expression changes and poorly reproduced spots, and to determine the number of replicates required to reliably detect a desired fold change in gene expression. ArrayStat can be integrated with the ArrayVision quantification program.

Reader Service No. 104

OmniUltra

ThermoHybaid www.thermoHybaid.com

Material gain

The injection-moulded polycarbonate used in this PCR plate is gamma irradiated for 100% sterility with no brittleness. The sealing characteristics of polycarbonate avoid the need for heated sealing solutions and the material is receptive to coatings for advanced protocols such as streptavidin for biotin-labelled PCR. The plate is designed for full robot compatibility. Samples are available from ThermoHybaid.

Reader Service No. 105

Proteomics catalogue

Pierce www.piercenet.com

"Proteomics Connection", published this year

This free 60-page catalogue features products used in cell lysis, total protein determination, protein staining, western blot detection and protein-protein interaction analysis. It contains a resources section listing proteomics-related books and recent journal articles.

Reader Service No. 106

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Polyclonal Antibodies & Custom Peptides

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen™
Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

DIVERSE SMALL MOLECULES READY FOR SCREENING



10K–50K
Compounds
in Plates

- Next Day Delivery*
- Competitively Priced

*Some Restrictions Apply

Chembridge Corporation • 800.964.6143
858.451.7400 • 858.451.7401 fax • E-mail: chem@chembridge.com

www.chembridge.com